International Journal of Digital Multimedia Broadcasting

Spectrum Sharing and Sensing for Future Broadband Networks: The Cognitive Radio Technology

Guest Editors: Massimiliano Laddomada, Hsiao Hwa Chen, Fred Daneshgaran, Marina Mondin, and Hamid Sadjadpour

Spectrum Sharing and Sensing for Future Broadband Networks: The Cognitive Radio Technology

International Journal of Digital Multimedia Broadcasting

Spectrum Sharing and Sensing for Future Broadband Networks: The Cognitive Radio Technology

Guest Editors: Massimiliano Laddomada, Hsiao Hwa Chen, Fred Daneshgaran, MarinaMondin, and Hamid Sadjadpour

Contents

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 898470, 2 pages
doi:10.1155/2010/898470

Editorial

Spectrum Sharing and Sensing for Future Broadband Networks: The Cognitive Radio Technology

Massimiliano Laddomada,[1] Hsiao Hwa Chen,[2] Fred Daneshgaran,[3] Marina Mondin,[4] and Hamid Sadjadpour[5]

[1] *Department of Electrical Engineering, Texas A&M University, Texarkana, TX 75505, USA*
[2] *National Cheng Kung University, Taiwan*
[3] *California State University, Los Angeles, USA*
[4] *Politecnico di Torino, Italy*
[5] *University of California, Santa Cruz, USA*

Correspondence should be addressed to Massimiliano Laddomada, mladdomada@tamut.edu

Received 26 June 2010; Accepted 29 June 2010

A recent measurement campaign on spectrum utilization promoted by the Federal Communication Commission throughout the US has revealed that significant portions of the electromagnetic spectrum are rarely exploited for considerable intervals of time, thus paving the way to a number of possibilities for increasing bandwidth utilization among multiple users in shared radio networks. The clue to improve spectrum utilization, as well as efficiency, comes from communications systems exploiting the so-called Cognitive Radio technology, which at the physical layer, as well as at upper-levels of the network, can adaptively and dynamically allow users to access radio resources by switching among portions of unused bandwidth in different intervals of time. The cognitive radio technology is still in its infancy, and many problems at a theoretical as well as practical level have to be solved before this technology may be fully exploited in next generation wireless networks.

This special issue is aimed at highlighting state-of-the-art techniques on the most recent research advances in cognitive radio networks. The first paper presents an optimization framework for collaborative spectrum sensing in terms of optimum decision fusion for hard and soft decision combining. One of the main conclusions of the work is that, for optimum fusion, the fusion centre must incorporate signal-to-noise ratio values of cognitive users and the channel conditions. A genetic algorithm based on weighted optimization strategy is also presented for the case of soft decision combining. In the second paper, a fast detection scheme consisting of multiple cognitive radios and a central control office is proposed. Specifically, each cognitive radio makes individual detection decision using the sequential probability ratio test combined with Neyman Pearson detection with respect to a specific observation window length. The third paper proposes the establishment of a simultaneous cognitive radio communication based on a subdistribution of power made over unselected subchannels that were discarded by the primary user through an initial optimal power allotment. The aim of the work is to show the possibility of introducing an opportunistic communication into a licensed transmission where the total power constraint is shared. The fourth paper proposes an algorithm for cooperative Dynamic Spectrum Access in Cognitive Radio networks utilizing Medium Access Control layer mechanisms for message exchange between secondary nodes in order to achieve interference mitigation. The proposed algorithm is applied in Filter Bank Multicarrier and Orthogonal Frequency Division Multiplexing systems. The fifth paper proposes an optimization technique based on genetic algorithms for optimal collaborative multiband sensing. Since Genetic programming performs a direct search of the optimal solution without approximations and solution domain restrictions, collaborative multiband sensing can be consistently optimized without limitations. The sixth paper investigates the application of a parallel compressive sensing

architecture for wideband spectrum sensing at sub-Nyquist rates by exploiting the current reduced frequency usage. The proposed architecture possesses attractive spurious frequency rejection schemes that are critical in the implementation of spectrum sensing systems. The seventh paper describes a cross-layer throughput optimization problem based on a modified signal-to-interference-plus-noise ratio. The objective is to maximize the minimum end-to-end flow throughput. The eighth paper presents a review of the state of the art in technology, regulation, and standardization of cognitive access to TV White Spaces. It examines the spectrum opportunity and commercial use cases associated with this form of secondary access. The ninth paper strives to provide a solution to the hidden-node problem for passive-listening receiver based on cooperation of multiple cognitive radios. The work considers a cooperative GPS-enabled cognitive network. Once the presence of primary users is detected, a localization algorithm is employed to first estimate the path loss model for the environment based on back propagation method and then to locate the position of primary user. In the tenth paper, two complementary algorithms based on information theoretic measures of statistical distribution divergence and information content are proposed. The first method is applicable to signals with periodic structures and is based on the analysis of Kullback-Leibler divergence. The second exploits information content analysis to detect unusual events. The last paper of this special issue focuses on a frequency-designed fractional delay FIR structure suitable for software radio applications. The design method is based on frequency optimization of a combination of modified Farrow and multirate structures.

The authors are grateful to the reviewers for their invaluable work and to the authors of the papers collected in this special issue.

Massimiliano Laddomada
Hsiao Hwa Chen
Fred Daneshgaran
Marina Mondin
Hamid Sadjadpour

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 424036, 20 pages
doi:10.1155/2010/424036

Research Article

Collaborative Spectrum Sensing Optimisation Algorithms for Cognitive Radio Networks

Kamran Arshad, Muhammad Ali Imran, and Klaus Moessner

Centre for Communication Systems Research, University of Surrey, Guildford GU2 7XH, UK

Correspondence should be addressed to Kamran Arshad, k.arshad@surrey.ac.uk

Received 28 October 2009; Accepted 17 March 2010

Academic Editor: Massimiliano Laddomada

The main challenge for a cognitive radio is to detect the existence of primary users reliably in order to minimise the interference to licensed communications. Hence, spectrum sensing is a most important requirement of a cognitive radio. However, due to the channel uncertainties, local observations are not reliable and collaboration among users is required. Selection of fusion rule at a common receiver has a direct impact on the overall spectrum sensing performance. In this paper, optimisation of collaborative spectrum sensing in terms of optimum decision fusion is studied for hard and soft decision combining. It is concluded that for optimum fusion, the fusion centre must incorporate signal-to-noise ratio values of cognitive users and the channel conditions. A genetic algorithm-based weighted optimisation strategy is presented for the case of soft decision combining. Numerical results show that the proposed optimised collaborative spectrum sensing schemes give better spectrum sensing performance.

1. Introduction

As numbers of wireless devices, innovative services, and number of mobile users continue to grow, more and more spectrum resources will be needed to guarantee desired Quality of Service. Mobile users want high-quality calls, streaming videos, and high-speed downloads, placing more and more stress on the limited radio spectrum available to the network operators. The radio spectrum spans around 300 billion frequencies; however, only a tiny fraction of frequencies can be used for commercial or personal radio communications; fundamental physical limits apply [1]. In the current spectrum regulatory framework, most frequency bands are exclusively allocated to the privileged users, often called Primary User (PU), which have all the rights to use the allocated bands. This approach protects PU's from any intersystem interference, but on the other hand, it yields highly inefficient use of the spectrum.

Measurements conducted by the Office of Communications (Ofcom) in UK and the Spectrum Policy Task Force (SPTF) in USA indicate that many chunks of the licensed spectrum are not used or only partially used, for significant periods of time [2, 3]. Spectrum occupancy measurements undertaken by Ofcom in Central London, at Heathrow airport and in some rural areas of the country, clearly show that there are significant portions of the radio spectrum which are not fully utilised in various geographical areas of the United Kingdom [4]. Similarly, in New York city maximum spectrum occupancy is reported as only 13.1% and downtown of Washington D.C. indicated spectrum occupancy of less than 35% of the radio spectrum below 3 GHz [5]. These studies clearly suggest that currently spectrum scarcity is mainly due to the inefficient use of spectrum rather than the physical shortage of spectrum. Particularly in UK, Olympic Games 2012 put extra pressure on Ofcom to plan the efficient use of radio spectrum to satisfy over 10 million spectators, around 15,000 participants and about 20,000 media personnel in the UK who will beam live pictures and commentary all around the world. Moreover, emerging as well as some existing operators are faced with the difficult task to gain access to the radio spectrum to operate their services [6]. In addition, access to a block of spectrum is very expensive as seen when the five operators were licensed for the 3rd generation mobile systems in the UK at a cost of around £22.5 billion [7]. More

recently (early 2009), the FCC spectrum auction in USA raised a record $19.9 billion dollars [8].

Cognitive Radio (CR) is widely regarded as the technology which will increase spectrum utilisation significantly in the next generation wireless communication systems by implementing opportunistic spectrum sharing. Spectrum sensing is one of the most critical functionalities in a cognitive radio network; it allows the unauthorised users, called Secondary Users (SUs), to detect unused portions of the spectrum called "spectrum holes" and opportunistically utilise these spectrum holes without causing harmful interference to the PU. The main goal of spectrum sensing is to obtain awareness about the spectrum usage and the existence of the PU in a certain geographical area at a particular period of time. In order to evaluate the performance of spectrum sensing, two metrics are of great interest: probability of detection and the probability of false alarm. Probability of detection, P_d, determines the level of interference-protection provided to the PU while probability of false alarm, P_f, indicates percentage of spectrum holes falsely declared as occupied [9]. In the context of opportunistic spectrum access, P_d must be higher than some predefined threshold while P_f should be lower than some desired criteria or as minimum as possible.

To enhance the performance of spectrum sensing, many techniques are available in the literature, and a brief survey has been recently published in [10]. In practice, CRs usually have no or limited knowledge about the primary signals; hence the optimal spectrum sensing technique is energy detection [11]. An energy detection approach for spectrum sensing at an individual CR has been assumed in this paper because of its simplicity, ease of implementation, and low computational complexity [9]. Moreover, the aim of this paper is to characterise gains achieved by collaboration of users without going into the details of complex local spectrum sensing schemes. The more sophisticated techniques like match filter detection or cyclostationary feature detection can be used for signal classification if more a priori knowledge about the structure of the primary signal is available [10]. However, performance of the energy detector is susceptible to noise power uncertainty [12]. Nevertheless, it has been shown that Collaborative Spectrum Sensing (CSS) is capable of delivering the desired detection performance under noise uncertainty for a large number of users [9]. However, energy detectors do not work efficiently for detecting spread spectrum signals [10]; spread spectrum signals are out of the scope of this paper.

1.1. Prior Work. The spectrum scarcity and spectrum underutilisation problem has stimulated a number of exciting activities in the technical, economic, and regulatory domains in searching for better spectrum management policies and techniques, for example, FCC opened up some analogue TV bands for unlicensed access [13]. However, spectrum sharing with PU must be done in a controlled way so that the PU operation in the particular frequency band is not disturbed. Furthermore, the IEEE standard 802.22 for unlicensed access to the TV bands is in its final stages of development [14]. Recently, Ofcom released Digital Dividend Review Statement (DDRS) which shows a radical shift in spectrum sharing policy in the UK and Ofcom is proposing to "allow license exempt use of interleaved spectrum for cognitive devices" [15]. Also, the European Commission (EU) paid much attention on dynamic spectrum management and the CR theme, and sponsored many FP5, FP6, and FP7 projects such as DRIVE [16], OverDRIVE [17], WINNER [18], E2R I/II [19], ORACLE [20], E3 [21], and "Radio Access and Spectrum" (RAS) cluster [22] tackling this issue. Similarly, several other projects outside Europe including the Defense Advanced Research Project Agency (DARPA)'s Next Generation program [23] and National Science Foundation program "NeTS-ProWiN" [24] show a significant momentum to shift spectrum access policy.

The cognition capability of a CR can make opportunistic spectrum access possible which can be implemented either by knowledge management mechanisms or by spectrum sensing functionality. A mobile network operator, for example, can equip the terminals with management mechanisms to select the most appropriate radio access technology of its heterogenous infrastructure [25, 26]. Concentrating on spectrum sensing, observations of a single CR are not always trustworthy because a CR may have good line of sight with the primary receiver but may not be able to see the primary transmitter due to shadowing or fading, known as "hidden node" problem. Collaboration has been proposed as a solution to the problems that arise due to such uncertainties in the channel. It has been shown many times in the literature that spectrum sensing performance can be greatly improved by CSS when a number of SUs share their sensing information; fusion of this information leads to a final decision about the existence of the PU. For an overview of recent advances in CSS, the reader can refer to [9, 27–32]. Existence of a large number of cognitive users creating multiple CRN's is highly probable in the future communication systems. However, the CSS mechanisms generate a large amount of traffic overhead since each SU needs to transmit its own decision; therefore collaboration of users needs to be refined and optimised [9].

Various techniques for the optimisation of CSS in terms of fusion rule [29], number of users [33], and thresholds [34] have been proposed. It has been argued in the literature that fusion schemes strongly impact on the spectrum sensing performance including probabilities of detection and false alarm [29]. In CSS, a CR can transmit either its local observations (soft decision) or a 1-bit decision (hard decision) to a common receiver, often called fusion centre. When hard decisions are combined at the fusion centre, the K-out-of-N fusion rule is normally used [33]. In the literature, there are some studies on the optimisation of the K-out-of-N rule to minimise total decision error probability [29] and to maximise the SUs throughput [35]; however, those algorithms were designed for a specific scenario of TV bands sharing in an AWGN channel. A fusion rule based on selected information for spectrum sensing is considered in [36], in which only users that have sufficient information send their 1-bit decision to the fusion centre and the fusion centre employs best fusion rule based on the received information.

A new fusion rule including "No decision" information from the cooperative nodes was proposed in [28].

The optimum fusion rule for combining soft decisions is Chair-Varshney rule which is based on log-likelihood ratio test [37]. Various other techniques for combining soft decisions are presented in [38]. However, most of the prior research work focuses on the case when SUs are far away from the primary transmitter and hence the same path loss or Signal-to-Noise Ratio (SNR) was assumed for all collaborating SUs [9, 29, 33]. The effects of different SNRs on detection performance are studied under AWGN channel conditions in [39]. Moreover, previous research highlighted CSS techniques which combine data from the CR nodes with equal weights and with perfect reporting channels [9, 40, 41], which is clearly not the case in realistic scenarios and might lead to misleading interpretation of results. The reporting channel for an ith user is defined as the channel between ith user and the fusion centre. Performance of CSS with noisy reporting channels was considered for the case of hard decision fusion in [31].

Collaborative spectrum sensing schemes with weighted user contributions have been recently proposed in [42, 43]. In [42] average signal power at an SU was exploited to assign weights to different collaborating cognitive nodes. In [43] a linear optimal strategy for CSS was presented and optimal weights for each SU in an AWGN channel were derived analytically. However, the shortcomings of existing literature in weighted CSS are in the fact that perfect reporting channels have been assumed instead of more realistic fading channels.

1.2. Major Contributions. In this paper, the optimisation of CSS is documented and optimum decision fusion is evaluated for hard and soft decision fusion at the fusion centre. Main contributions of this paper are summarised as follows.

(i) Hard decision fusion is attractive because of lower communication overhead over the reporting channels. In this paper, the problem of hard decision fusion at the fusion centre is addressed and answers this simple question: for optimal fusion does the fusion centre only need 1-bit decision? Different scenarios are considered with users close to the primary transmitter have the different SNR values. It is concluded that in order to achieve optimum spectrum sensing performance, the fusion centre must have SNR information for each CR and channel conditions along with their 1-bit decisions.

(ii) Maximum diversity in CSS is achieved when all collaborating users experience identical and independent fading or shadowing effects, which is not possible in reality if users are too close to each other. Multipath fading can be assumed to be independent from one user to another but shadowing is normally correlated over large distances. Thus, secondary users in close vicinity of each other make similar measurements and this limits the collaboration gains. In this paper, correlated log-normal shadowing is considered among collaborative users and it is shown that correlated shadowing has direct impact on the optimal fusion rule at the fusion centre.

(iii) Genetic Algorithm- (GA-) based weighted collaborative spectrum sensing strategy is proposed in this paper to combat the effects of channel and enhance spectrum sensing performance. The proposed optimum spectrum sensing framework is based on a model that is realistic and also takes into account both channels, that is, channel between PU and SUs as well as the reporting channels. It is shown in this paper that imperfect reporting channel and different SU SNR values have direct impact on the performance of CSS. Secondary users transmit their soft decisions to the fusion centre and a global decision is made at the fusion centre which is based on a weighted combination of the local test statistics from individual SUs. The weight of each SU is indicative of its contribution to the final decision making. For example, if an SU has a high SNR signal and also has a good reporting channel (higher reporting channel gain), then it is assigned a larger contributing weight. The optimum CSS problem is formulated as a nonlinear optimisation problem in this paper. For a given probability of false alarm and channel conditions, optimal weights are chosen in such a way that it maximises global probability of detection at the fusion centre. With a realistic fading channel it is hard to derive an analytical expression for the optimum weights and hence a GA-based solution is proposed.

1.3. Organisation of the Paper. The remainder of this paper is organised as follows. In Section 2 the system model is briefly introduced and the use cases are defined. Section 3 discusses local spectrum sensing under channel fading conditions and its limitations. Section 4 briefly explains CSS and decision fusion techniques for both HDC and SDC, considered in this paper. Section 5 proposes a framework for optimisation of fusion rules for HDC. In order to achieve optimum spectrum sensing performance, GA is used to calculate the weights for each collaborative user in Section 6. Finally Section 7 concludes the paper.

2. System Model for Cognitive Radio Network

Consider a cognitive radio network, with M cognitive users (indexed by $i \in \{1, 2, \ldots, M\}$), and a fusion centre to sense a portion of the spectrum of bandwidth "W" in order to detect the existence of the PU, as shown in Figure 1. Assume that each CR is equipped with an energy detector and is able to perform local spectrum sensing independently. Each CR makes its own observation based on the received signal, that is, noise only or signal plus noise. Hence, the spectrum sensing problem can be considered as a binary hypothesis

testing problem with two possible hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ defined as [38]

$$x_i(t) = \begin{cases} n_i(t), & \mathcal{H}_0, \\ h_i s(t) + n_i(t), & \mathcal{H}_1, \end{cases} \tag{1}$$

where $s(t)$ is the PU signal and is assumed to be an identical and independent random process (i.i.d.) with zero mean and variance σ_s^2. For the ith SU, the receiver noise is modelled as $n_i(t)$ which is also assumed to be an i.i.d. random process with zero mean and variance σ_n^2 and h_i is the complex gain of the channel between the PU and the ith SU. Further, it is assumed that $s(t)$ and $n_i(t)$ are independent of each other. The power transmitted by the PU is received at the SU and the ratio of received power to the power of noise at the SU is defined as the SNR at the SU energy detector. The received SNR at the ith SU can be more precisely defined as

$$\gamma_i \triangleq \frac{\mathrm{E}\left[|h_i|^2\right]\sigma_s^2}{\sigma_n^2}. \tag{2}$$

System model and use cases for considered scenarios are shown in Figure 1. Two use cases are assumed in this paper. Use Case 1 refers to the case when PU transmitter is far away from the CRN and hence same SNR can be assumed for all SUs. In use Case 2, the PU is not far away from the M SUs and each user has a different value of SNR depending on its distance from the PU and its channel conditions.

3. Local Spectrum Sensing

The performance of a given spectrum sensing scheme is fundamentally limited by the radio propagation channel. Typically, the effects of a radio channel can be divided into three main parts: path loss, small-scale fading, and large-scale fading (shadowing) [44]. Path loss effects are incorporated in the received SNR at a cognitive radio terminal. Small-scale fading causes rapid, random variations in the signal strength at the CR receiver and is modelled by Rayleigh fading in this paper. Shadowing is the slow variation of received signal power as the cognitive radio moves in and out of the shadow of large structures like mountains, buildings, and so forth. Shadowing is often modelled as a log-normal distributed random process that varies around a local mean given by the path loss and with the standard deviation σ_{dB} which depends on the environment [45].

3.1. AWGN Channel. In energy detection-based spectrum sensing, the received radio frequency energy in the considered channel or frequency band W is measured over a time interval T to determine whether the PU signal $s(t)$ is present. Assume that the time bandwidth product is always an integer and is denoted by $N = TW$. Test statistic u_i calculated by an ith user is given as

$$u_i = \sum_{k=1}^{N} \left| x_i\left(\frac{k}{W}\right) \right|^2. \tag{3}$$

u_i is compared with a predefined threshold λ_i to get the local decision:

$$u_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_i. \tag{4}$$

The binary decision is given by D_i; $D_i = 1$ when $u_i > \lambda_i$ and 0 otherwise. u_i is the sum of squares of N Gaussian random variables and it is well known that the sum of squares of Gaussian variables follows a chi-square distribution [46]. Thus u_i follows a central chi-square distribution with $2N$ degrees of freedom under hypothesis $\mathcal{H}_0$ and a noncentral chi-square distribution with $2N$ degrees of freedom and non-centrality parameter of $2N\gamma_i$ under hypothesis $\mathcal{H}_1$. Therefore, the probability density function (pdf) of random variable U_i under the two hypotheses can be written as

$$f_{U_i}(u) = \begin{cases} \dfrac{u^{N-1}e^{-u/2}}{2^N\Gamma(N)}, & \mathcal{H}_0, \\ \dfrac{1}{2}\left(\dfrac{u}{2N\gamma_i}\right)^{(N-1)/2} e^{-(u+2N\gamma_i)/2} I_{N-1}\left(\sqrt{2Nu\gamma_i}\right), & \mathcal{H}_1, \end{cases} \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function and $I_{N-1}(\cdot)$ is the modified Bessel function of the first kind. For an ith user probability of false alarm, $\Pr(\mathcal{H}_1 \mid \mathcal{H}_0)$, and detection, $\Pr(\mathcal{H}_1 \mid \mathcal{H}_1)$ can be derived from (5) and is given as

$$P_f^i = \Pr\{U_i > \lambda_i \mid \mathcal{H}_0\} = \frac{\Gamma(N, \lambda_i/2)}{\Gamma(N)}, \tag{6}$$

$$P_d^i = \Pr\{U_i > \lambda_i \mid \mathcal{H}_1\} = Q_N\left(\sqrt{2N\gamma_i}, \sqrt{\lambda_i}\right), \tag{7}$$

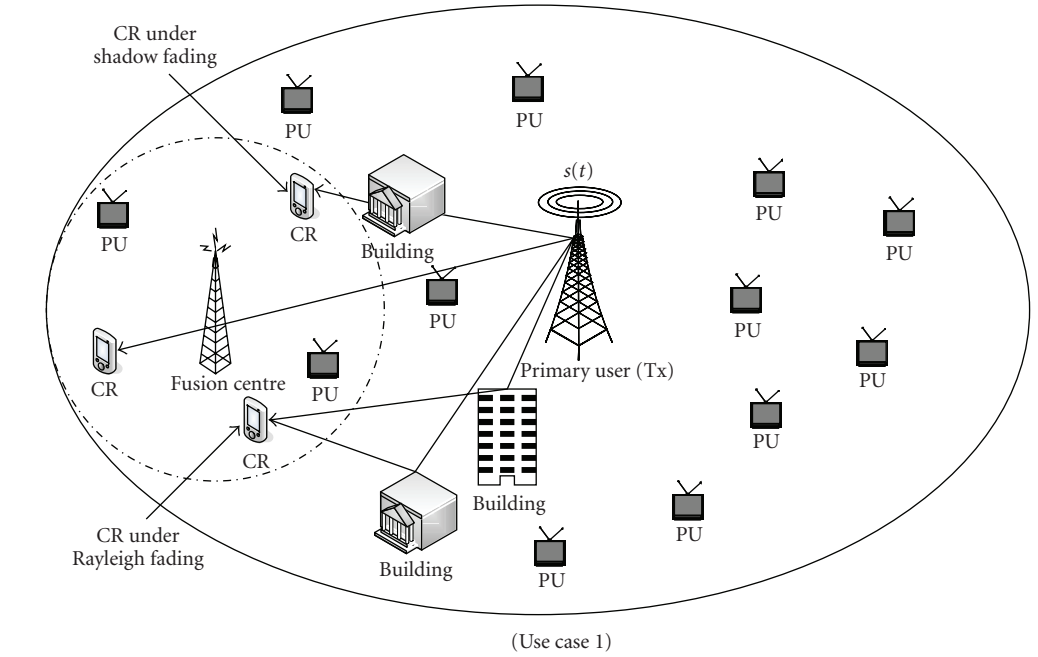
where $\Gamma(a, x)$ is incomplete gamma function and $Q_N(a, b)$ is the generalised Marcum Q-function. Detailed derivations of P_f^i and P_d^i are given in Appendices A and B.

For the purpose of simplifying (5) an approximate model for energy detection-based spectrum sensing observations can be built. It has been shown in [47] that the approximated model converges faster and has lower approximation error when N is asymptotically large. So when N tends towards infinity (practically when $N \geq 10$ [46]), the chi-square distribution defined in (5) converges to a normal distribution, that is,

$$U_i \sim \begin{cases} \mathcal{N}\left(N\sigma_i^2, 2N\sigma_i^4\right), & \mathcal{H}_0, \\ \mathcal{N}\left((N+\gamma_i)\sigma_i^2, 2(N+2\gamma_i)\sigma_i^4\right), & \mathcal{H}_1. \end{cases} \tag{8}$$

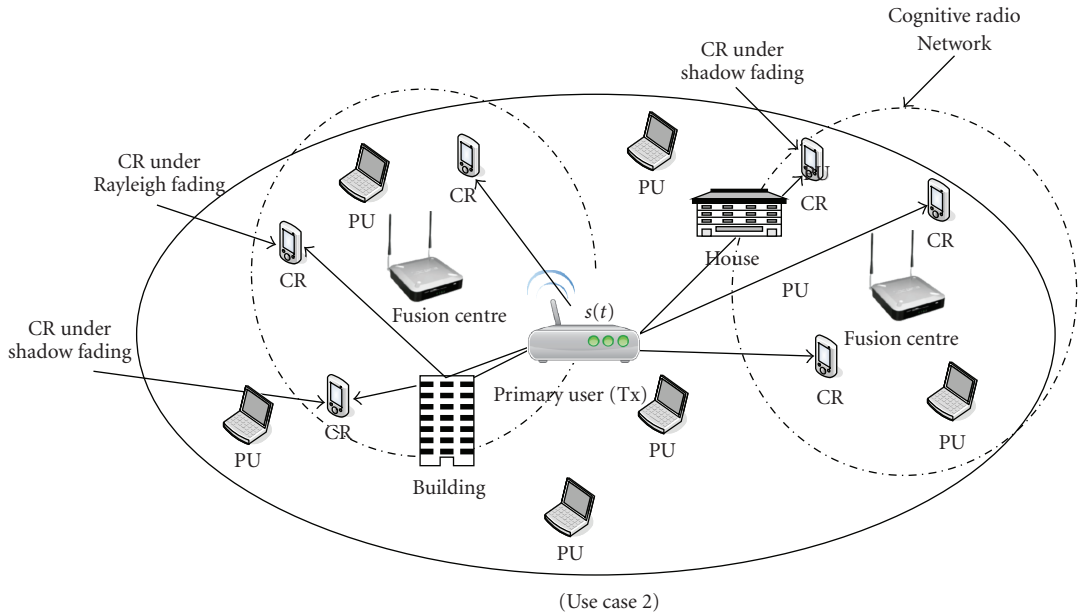
Similarly, P_f^i and P_d^i defined in (6) and (7) can be approximated as

$$P_f^i = Q\left(\frac{\lambda_i - \mathbb{E}[U_i \mid \mathcal{H}_0]}{\sqrt{\mathrm{Var}[U_i \mid \mathcal{H}_0]}}\right) = Q\left(\frac{\lambda_i - N\sigma_i^2}{\sqrt{2N}\sigma_i^2}\right),$$



(Use case 1)

(a)



(Use case 2)

(b)

Figure 1: Use cases considered in paper.

$$P_d^i = Q\left(\frac{\lambda_i - \mathbb{E}[U_i \mid \mathcal{H}_0]}{\sqrt{\mathrm{Var}[U_i \mid \mathcal{H}_0]}}\right) = Q\left(\frac{\lambda_i - (N+\gamma_i)\sigma_i^2}{\sqrt{2(N+2\gamma_i)\sigma_i^2}}\right), \quad (9)$$

where $\mathbb{E}[\cdot]$ and $\mathrm{Var}\,g[\cdot]$ denote expectation and variance operations, respectively.

3.2. Spectrum Sensing in Fading Channels. When the SU is in a fading channel, the channel gain h_i for an ith user is varying due to the fading and P_d^i becomes conditional probability dependent on instantaneous SNR γ_i. As expected, P_f^i is independent of γ_i and remains static. Average probability of detection can be obtained by averaging instantaneous P_d^i over fading statistics, where the pdf of received SNR is $f_\gamma(x)$:

$$P_{d,\text{fading}}^i = \int_\gamma P_d^i(x) f_\gamma(x) dx. \quad (10)$$

When the channel is Rayleigh faded, then γ_i is exponentially distributed with $\overline{\gamma}_i$ as its mean value. Similarly when channel is shadow faded, then γ_i is log-normally distributed with

mean $\overline{\gamma}_i$ and characterised by dB-spread of shadowing σ_{dB} [38]. Hence,

$$f_{\gamma_i}(x) = \begin{cases} \dfrac{1}{x}\exp\left(-\dfrac{x}{\overline{\gamma}_i}\right); & \text{Rayleigh Fading,} \\ \dfrac{\xi}{x\sigma_{\text{dB}}\sqrt{(2\pi)}}\exp\left(-\dfrac{10\log_{10}(x)-\mu_{x_{\text{dB}}}}{2\sigma_{\text{dB}}^2}\right); & \text{Shadow Fading,} \end{cases} \tag{11}$$

where $\xi = 10/\ln(10)$ and $\mu_{x_{\text{dB}}}$ is the mean of $x_{\text{dB}} = 10\log(x)$. The conversion from linear mean to the log mean (in dB) can be derived as [45]

$$\mu_{x_{\text{dB}}} = 10\log_{10}(\overline{\gamma_i}) - \frac{\sigma_{\text{dB}}^2}{2\xi}. \tag{12}$$

Substituting (7) and (11) in (10), for Rayleigh fading channel average probability of detection for the ith user can be calculated by [38]

$$\begin{aligned} P_{d,\text{rayl}}^i &= \int_\gamma \frac{1}{x} Q_N\left(\sqrt{2Nx}, \sqrt{\lambda_i}\right)\exp\left(-\frac{x}{\overline{\gamma}_i}\right)dx \\ &= e^{-\lambda_i/2}\sum_{n=0}^{N-2}\frac{1}{n!}\left(\frac{\lambda_i}{2}\right)^N + \left(\frac{1+\overline{\gamma_i}}{\overline{\gamma_i}}\right)^{N-1} \\ &\quad\times\left[e^{-\lambda_i/2(1+\overline{\gamma}_i)} - e^{-\lambda_i/2}\sum_{n=0}^{N-2}\frac{1}{n!}\frac{\lambda_i\overline{\gamma_i}}{2(1+\overline{\gamma_i})}\right]. \end{aligned} \tag{13}$$

For shadow fading, close form solution of (10) is not known and a numerical solution is required:

$$\begin{aligned} P_{d,\text{shadow}}^i &= \int_\gamma Q_N\left(\sqrt{2Nx}, \sqrt{\lambda_i}\right)\frac{1}{x\sigma_{\text{dB}}\sqrt{2\pi}} \\ &\quad\times\exp\left(-\frac{10\log(x)-\mu_{x_{\text{dB}}}}{2\sigma_{\text{dB}}^2}\right)dx \\ &= \frac{1}{\sigma_{\text{dB}}\sqrt{2\pi}}\sum_{x=x_0}^{x_f} Q_N\left(\sqrt{2Nx}, \sqrt{\lambda_i}\right) \\ &\quad\times\exp\left(-\frac{10\log(x)-\mu_{x_{\text{dB}}}}{2\sigma_{\text{dB}}^2}\right)\frac{\Delta x}{x}, \end{aligned} \tag{14}$$

where Δx and x_f are chosen as to minimise numerical approximation error.

3.3. Numerical Evidence. The performance of local spectrum sensing is evaluated using theoretical results as well as Monte Carlo simulations by plotting complementary Receiver Operating Characteristics (ROC) curves (plot of $P_m = 1 - P_d$ versus P_f). In Monte Carlo simulations, probability of false alarm and miss detection is calculated by comparing sensing observations with a predefined threshold, and results are obtained by simulations over 1,000,000 noise realisations. It is assumed that N is an integer value and set to be 5.

Figure 2 shows the ROC curves for local spectrum sensing in AWGN, Rayleigh fading, and Shadowing for different values of σ_{dB}. Spectrum sensing results for AWGN channel are provided for comparison and simulation results are validated by comparing with analytical results. It is clear from Figure 2 that both Rayleigh and shadow fading degrades the performance of spectrum sensing. For example, in Rayleigh fading channel, in order to achieve $P_m < 10^{-1}$ where $P_m = 1 - P_d$, we need $P_f > 0.4$ which results in poor spectrum utilisation and vice versa. Similarly, it can be seen from Figure 2 that local spectrum sensing is more difficult in shadow fading and with increase in shadowing (or σ_{dB}) detector performance further degrades.

Another important metric to characterise spectrum sensing performance is the minimum detected SNR. This metric is defined as the lowest SNR that a sensing algorithm is able to detect with reliability of P_f and P_d for a given PU signal, propagation conditions and observation time. Figure 3 plots the minimum detectable SNR by a CR under different channel conditions for a targeted $P_f = 10^{-1}$. It is clear from Figure 3 that shadowing affects detector performance more than Rayleigh fading. In order to achieve $P_f = 10^{-1}$ in given scenario, the required SNR is around 10 dB while for the lower values of $\overline{\gamma}$ this is not possible as shown in Figure 3.

4. Collaborative Spectrum Sensing

Section 3 shows that local spectrum sensing has some limitations and it is hard to detect signals of low SNR for desired performance. Among many other challenges (e.g., see [48]) one of the most important challenges for the implementation of CRN is the hidden node problem, when a CR is shadowed or in a deep fade [41]. To address these problems multiple CRs can collaborate with each other in order to make a global decision about the existence of the PU. It has been shown by previous research that CSS can improve detection performance in the fading channels; for example, see [9] and references therein. In CSS, every SU performs its own spectrum sensing measurements and can also make a local decision on whether the PU is present or absent. All of the SUs forward their soft (local measurement) or hard (1-bit) decision to a common receiver, often called fusion centre or a band manager. Fusion centre may be centralised or distributed; in centralised CSS all the SUs send their decisions to the fusion centre, which may be an Access Point (AP) in wireless LAN or a CR base station in a cellular system, while, in distributed CSS, all the SUs may behave as a fusion centre and receive sensing information from the neighboring nodes. In both cases, fusion centre fuse collected decisions and make a final decision to declare the presence (or otherwise) of primary users in observed frequency band. The results presented in [40, 49] show that SDC outperforms HDC in terms of probability of miss detection. While HDC outperforms SDC when the

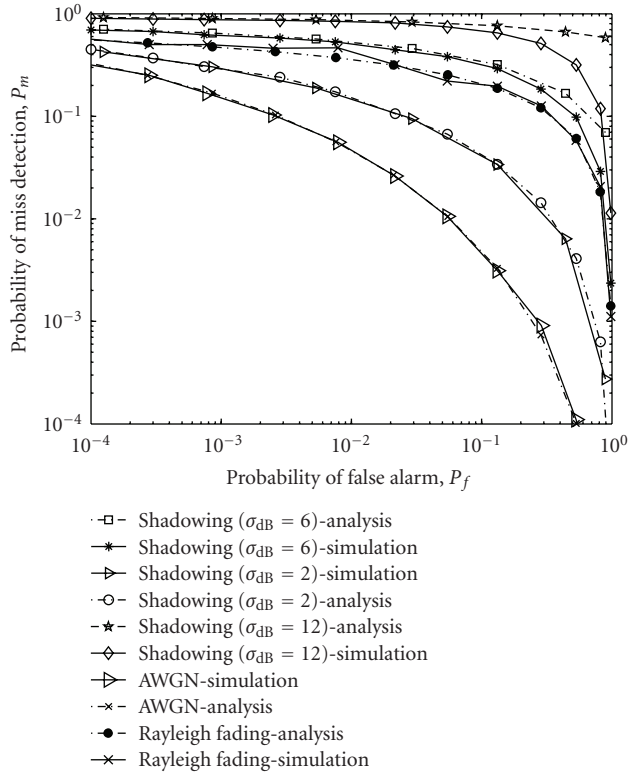


Figure 2: Receiver operating characteristics for local spectrum sensing in Rayleigh and Shadow Fading channels with $\gamma = 5$ dB, $N = 5$.

number of collaborative users is large [41] and further HDC needs a low-bandwidth control channel. In CSS, sharing information within CRN and combining result from various measurements is a challenging task, which is the main scope of this paper.

4.1. Hard Decision Combining. In HDC, fusion centre collects binary decisions from the individual SUs, identifies the available spectrum, and then broadcasts this information to the other SUs. The optimal decision fusion is based on Neyman-Pearson criterion by comparing Likelihood Ratio with the threshold vector as

$$\frac{f(\mathbf{D} \mid \mathcal{H}_1)}{f(\mathbf{D} \mid \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \underline{\lambda}, \tag{15}$$

where $\mathbf{D} = [D_1, D_2, \ldots, D_M]^T$ denotes binary decisions from M SUs and $D_i \in \{0, 1\}$, $\underline{\lambda}$ is the optimal threshold vector and $f(\mathbf{D} \mid \mathcal{H}_0)$, and $f(\mathbf{D} \mid \mathcal{H}_1)$ represents the probability density functions of $\mathbf{D}$ under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Mathematical analysis using Neyman-Pearson criterion is mathematically untractable especially if the local measurements are correlated and hence sub optimal solutions are always preferable [50].

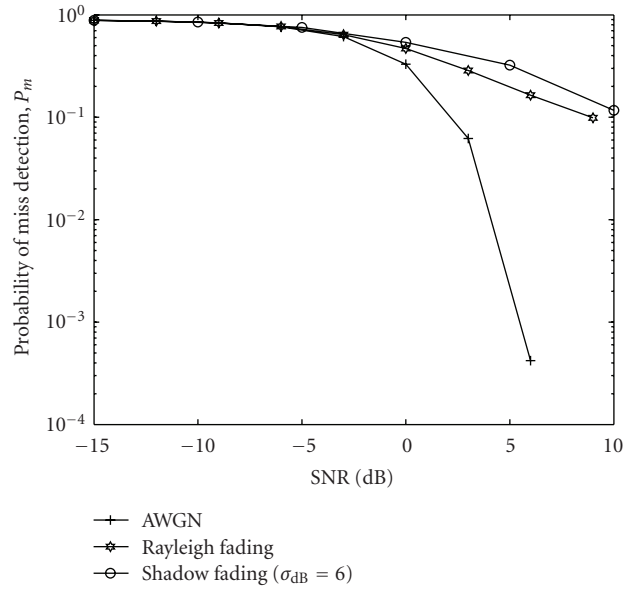


Figure 3: Probability of miss detection versus minimum detected SNR in shadow fading for $P_f = 10^{-1}$, $\gamma_{\text{dB}} = 5$, $N = 5$.

There are many other ways to combine or fuse hard decisions based on counting rules; most commonly used in the literature are *OR*, *AND* and in general K-out-of-M fusion rule [36, 42, 51]. In *AND* all CRs should declare $\mathcal{H}_1$ in order to make a global decision that PU is present while in *OR* rule, fusion centre declares $\mathcal{H}_1$ if any of the received decision is $\mathcal{H}_1$. At the fusion centre, all D_i's are fused together according to the following fusion rule [9]:

$$y_c = \begin{cases} \sum_{i=1}^{M} D_i \geq K, & \mathcal{H}_1, \\ \sum_{i=1}^{M} D_i \leq K, & \mathcal{H}_0. \end{cases} \tag{16}$$

It can be seen from (16) that the *OR* corresponds to the case when $K = 1$ while for *AND* rule $K = M$.

It has been reported that for many cases of practical interest, the *OR* fusion rule delivers better performance [9]. In order to demonstrate improvement in spectrum sensing performance by collaboration of SUs *OR* fusion rule is used at the fusion centre in this section. Figures 4 and 5 show ROC curves for use Case 1 (as shown in Figure 1) with different number of CRs under i.i.d. log-normal shadowing with $\overline{\gamma}_1 = \overline{\gamma}_2 = \cdots = \overline{\gamma}_i = 5$ dB and $N = 5$. In these results, AWGN curves for single users are shown for comparison. As seen in Figures 4 and 5 CSS mitigates the effects of shadow fading effectively. It can also be seen in Figure 4 that by incorporating more and more users performance even better than in the AWGN scenario can be achieved. This stems from the fact that with more number of SUs there are more chances that a single user has its instantaneous SNR above average.

As stated in Section 3 another important parameter to analyse performance of a detection algorithm is minimum

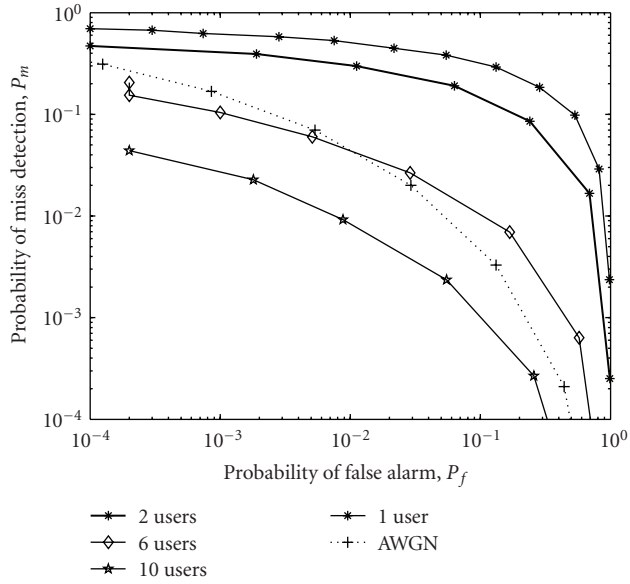


Figure 4: Receiver operating characteristics for collaborative spectrum sensing under shadow fading, $\sigma_{\text{dB}} = 6$, $N = 5$.

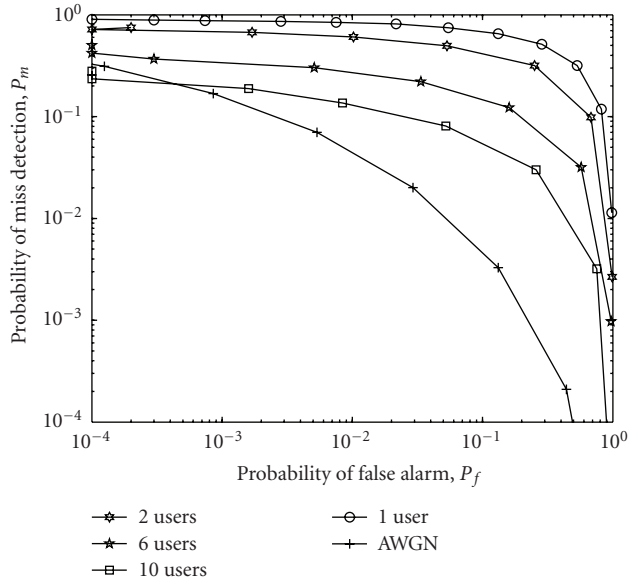


Figure 5: Receiver operating characteristics for collaborative spectrum sensing under shadowing, $\sigma_{\text{dB}} = 10$, $N = 5$.

detected SNR. A good detection scheme must be able to detect signals of low SNR, and in Section 3 it has been shown that shadowing affects detected SNR by a user. Figure 6 shows that by incorporating a large number of users it is possible to achieve the desired performance even at low SNR levels. By comparing Figures 3 and 6 it can be seen that under shadow fading ($\sigma_{\text{dB}} = 6$) and for desired performance, for example, $P_f = 10^{-1}$ and $P_d = 10^{-1}$ local sensing requires received signal of at least 10 dB while collaboration of 20 users can detect signal of SNR as low as -15 dB.

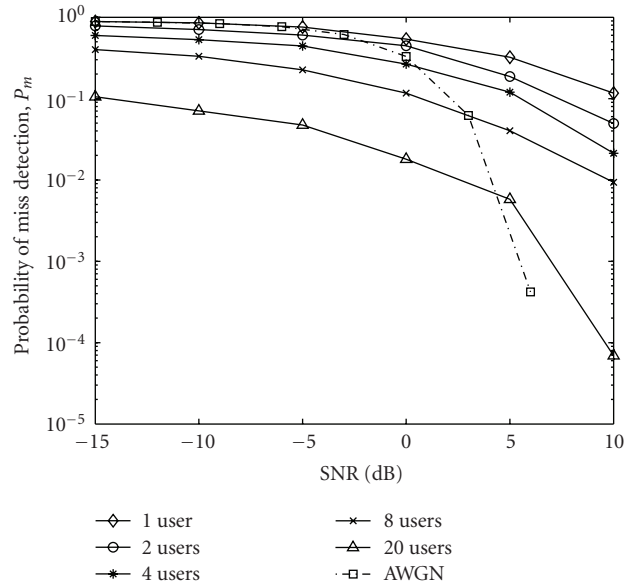


Figure 6: Probability of miss detection versus minimum detected SNR in shadow fading, $P_f = 10^{-1}$, $\sigma_{\text{dB}} = 6$, $\gamma = 5$ dB, $N = 5$.

4.2. Soft Decision Combining. In order to simplify the analysis with fusion of soft decisions, it has been assumed that the value of N is large. With this assumption the summary statistics at local secondary nodes $\mathbf{U}$ (as defined in (8)) can be considered, which is transmitted to the fusion centre through the reporting channels. In this paper realistic noisy reporting channels with variable channel gains are considered. A system model is shown in Figure 13.

4.2.1. Equal Gain Combining. Statistics of local observations for an ith SU after passing through the channel of gain g_i and noise $n_i \sim \mathcal{N}(0, \delta_i^2)$ is

$$y_i \sim \begin{cases} \mathcal{N}\left(Ng_i\sigma_i^2, 2Ng_i^2\sigma_i^4 + \delta_i^2\right), & \mathcal{H}_0, \\ \mathcal{N}\left(\left(N + \overline{\gamma}_i\right)g_i\sigma_i^2, 2\left(N + 2\overline{\gamma}_i\right)g_i^2\sigma_i^4 + \delta_i^2\right), & \mathcal{H}_1, \end{cases} \tag{17}$$

where δ_i^2 is the noise variance of the ith reporting channel. For the soft decision fusion scheme, fusion centre decides between $\mathcal{H}_0$ and $\mathcal{H}_1$ by comparing sum of individual observations y_c with a global threshold λ_c:

$$y_c = \sum_{i=1}^{M} y_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda_c. \tag{18}$$

4.2.2. Weighted Combining. In weighted combining, global test statistics is calculated at the fusion centre by assigning weights w_i to the received observation from an ith user y_i by

$$y_c = \sum_{i=1}^{M} w_i \cdot y_i = \mathbf{w}^T\mathbf{y}, \tag{19}$$

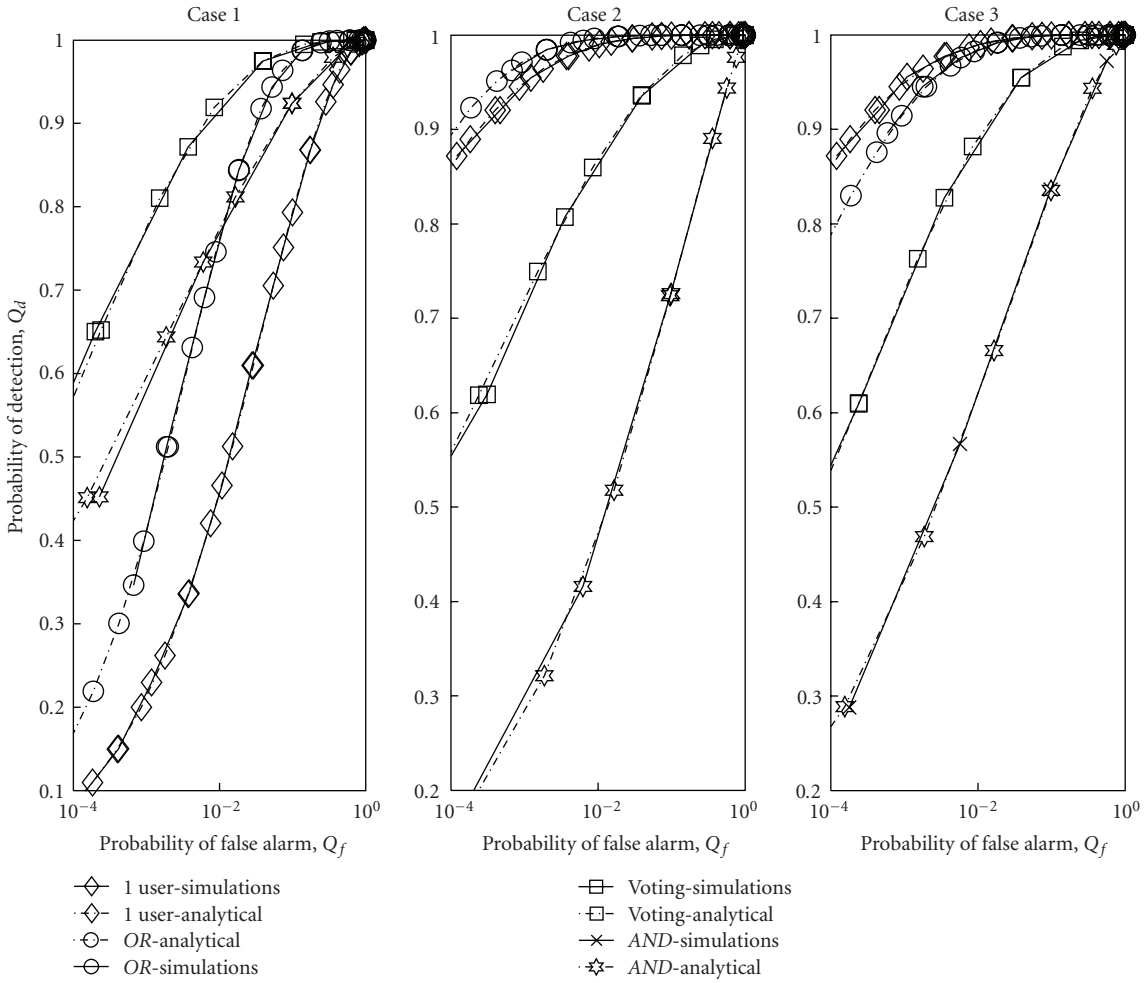


Figure 7: Receiver operating characteristics for 5 collaborating users in AWGN Channel in 3 cases: Case 1 (all users have similar SNR), Case 2 (half of the users have high SNR), and Case 3 (only one user has high SNR).

where $\mathbf{w} = [w_1, w_2, \ldots, w_M]^T \in \mathbb{R}^{M\times 1}$ and the received decision vector at the fusion centre is defined as $\mathbf{y} = [y_1, y_2, \ldots, y_M]^T \in \mathbb{R}^{M\times 1}$. Weight vector $\mathbf{w}$ at the fusion centre satisfies $\sum_{i=1}^{M} w_i = 1$. From (17) and (19) the distribution of y_c is given as

$$y_c \sim \begin{cases} \mathcal{N}\left(\sum_{i=1}^{M} N g_i \sigma_i^2 w_i, \sum_{i=1}^{M} \left(2N g_i^2 \sigma_i^4 w_i^2 + \delta_i^2 w_i^2\right)\right); & \mathcal{H}_0 \\ \mathcal{N}\left(\sum_{i=1}^{M} \left(\left(N+\overline{\gamma}_i\right) g_i \sigma_i^2 w_i\right), \right. \\ \qquad \left. \sum_{i=1}^{M} \left(2\left(N+2\overline{\gamma}_i\right) g_i^2 \sigma_i^4 w_i^2 + \delta_i^2 w_i^2\right)\right); & \mathcal{H}_1. \end{cases} \tag{20}$$

Assume $\mathbf{h} = [h_1, h_2, \ldots, h_M]^T \in \mathbb{R}^{M\times 1}$, $\mathbf{g} = [g_1, g_2, \ldots, g_M]^T \in \mathbb{R}^{M\times 1}$, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_M]^T \in \mathbb{R}^{M\times 1}$, $\boldsymbol{\sigma} = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2]^T \in \mathbb{R}^{M\times 1}$, and $\boldsymbol{\delta} = [\delta_1^2, \delta_2^2, \ldots, \delta_M^2]^T \in \mathbb{R}^{M\times 1}$. Furthermore, defined matrices $\boldsymbol{\Sigma}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$, and $\mathbf{G}$ that all belong to $\mathbb{R}^{M\times M}$ represent the diagonal matrices formed by placing the vectors $\boldsymbol{\sigma}$, $\boldsymbol{\delta}$, $\overline{\boldsymbol{\gamma}}$, and $\mathbf{g}$ on the diagonal, respectively. The statistics of y_c under $\mathcal{H}_0$ and $\mathcal{H}_1$ can be written as

$$\begin{aligned} \mathrm{E}[y_c \mid \mathcal{H}_0] &= N\mathbf{g}^T\boldsymbol{\Sigma}\mathbf{w}, \\ \mathrm{Var}[y_c \mid \mathcal{H}_0] &= \mathbf{w}^T\left[2N\mathbf{G}^2\boldsymbol{\Sigma}^2 + \boldsymbol{\Delta}\right]\mathbf{w}, \\ \mathrm{E}[y_c \mid \mathcal{H}_1] &= \mathbf{g}^T((N\mathbf{I} + \boldsymbol{\Gamma}) \odot \boldsymbol{\sigma})\mathbf{w}, \\ \mathrm{Var}[y_c \mid \mathcal{H}_1] &= \mathbf{w}^T\left[2(N\mathbf{I} + 2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2 + \boldsymbol{\Delta}\right]\mathbf{w}. \end{aligned} \tag{21}$$

To make a decision on the presence of a primary transmitter, the global decision statistic y_c as defined in (21) is compared with a threshold λ_c. Global probability of false

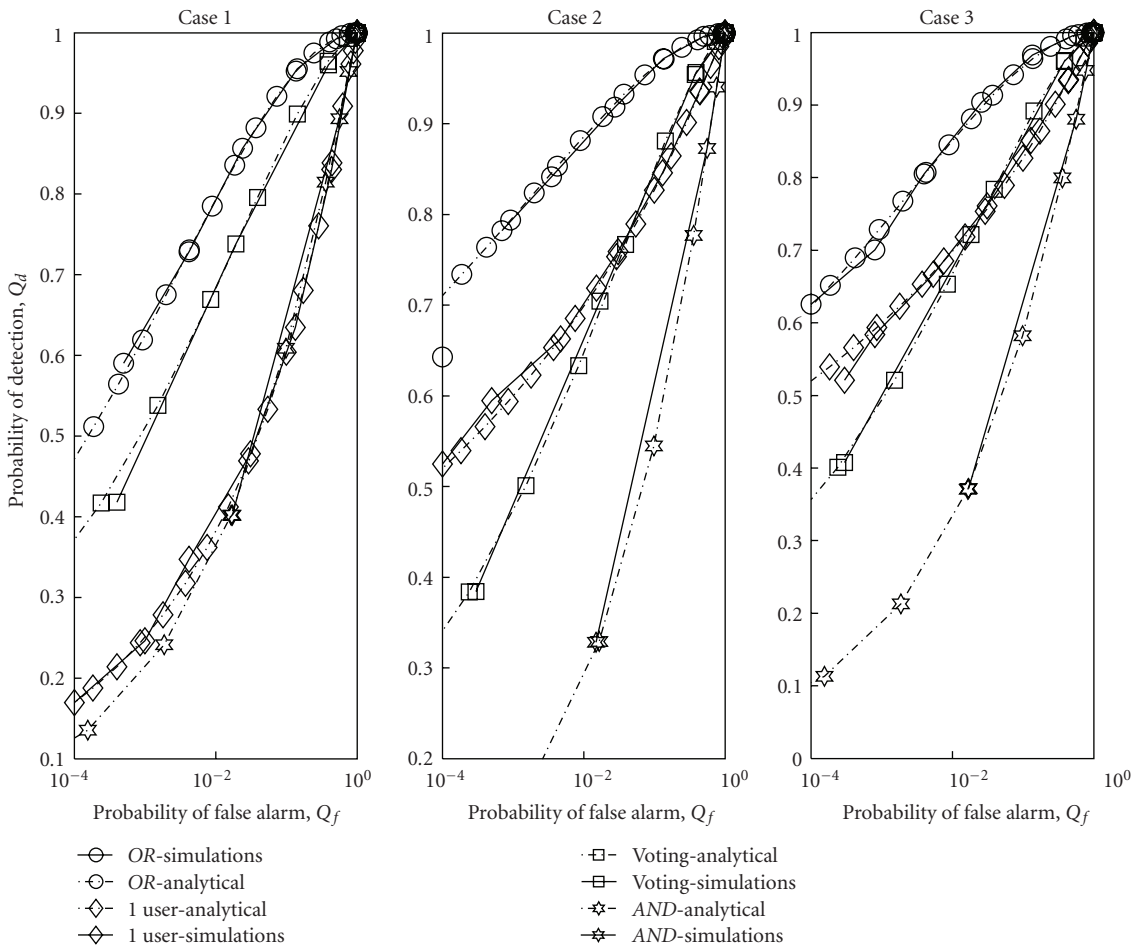


Figure 8: Receiver operating characteristics for 5 collaborating users in Rayleigh fading Channel in 3 cases: Case 1 (all users have similar SNR), Case 2 (half of the users have high SNR), and Case 3 (only one user has high SNR).

alarm and detection at the fusion centre, as denoted by Q_f and Q_d, are given as

$$Q_f = Q\left(\frac{\lambda_c - N\mathbf{g}^T\boldsymbol{\Sigma}\mathbf{w}}{\sqrt{\mathbf{w}^T\left[2N\mathbf{G}^2\boldsymbol{\Sigma}^2 + \boldsymbol{\Delta}\right]\mathbf{w}}}\right),$$
$$Q_d = Q\left(\frac{\lambda_c - \mathbf{g}^T((N\mathbf{I} + \boldsymbol{\Gamma}) \odot \boldsymbol{\sigma})\mathbf{w}}{\sqrt{\mathbf{w}^T\left[2(N\mathbf{I} + 2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2 + \boldsymbol{\Delta}\right]\mathbf{w}}}\right), \quad (22)$$

where $Q(\cdot)$ is the tail probability of the normalised Gaussian distribution.

5. Optimised User Collaboration Scheme for HDC

Section 4.1 shows that collaboration of SUs improves spectrum sensing performance by utilising space diversity of users. In this section, the problem of hard decision fusion at the fusion centre is considered in the presence of i.i.d. and spatially correlated shadowing. In the past, emphasis was given to collaborative spectrum sensing when all users have same received SNR; however, in this section, a scenario where users have different γ_i with AWGN and log-normal shadowing is considered. Three different cases in use Case 2 are considered here which represents three different scenarios depending on the location of PU and SUs. Case 1 refers to a scenario in which all the SUs are relatively close to each other and hence having similar values of SNR. Case 2 depicts the situation when half of the collaborating users have high SNR values while in Case 3 only one use has a high SNR value as compared to other collaborating SUs.

Different decision fusion schemes at the fusion centre including *OR*, *AND*, *Voting*, and *1-user* cases are considered. In *Voting*-based decision fusion scheme all SUs vote and fusion centre declare an opportunity if the majority of the collaborative SUs declare an opportunity. In *1-user* case although fusion centre receives information from all users, it uses only one user information in order to make a global decision.

5.1. Independent and Identically Distributed Shadowing

5.1.1. Mathematical Formulation. The global probability of detection Q_d and probability of false alarm Q_f at the fusion

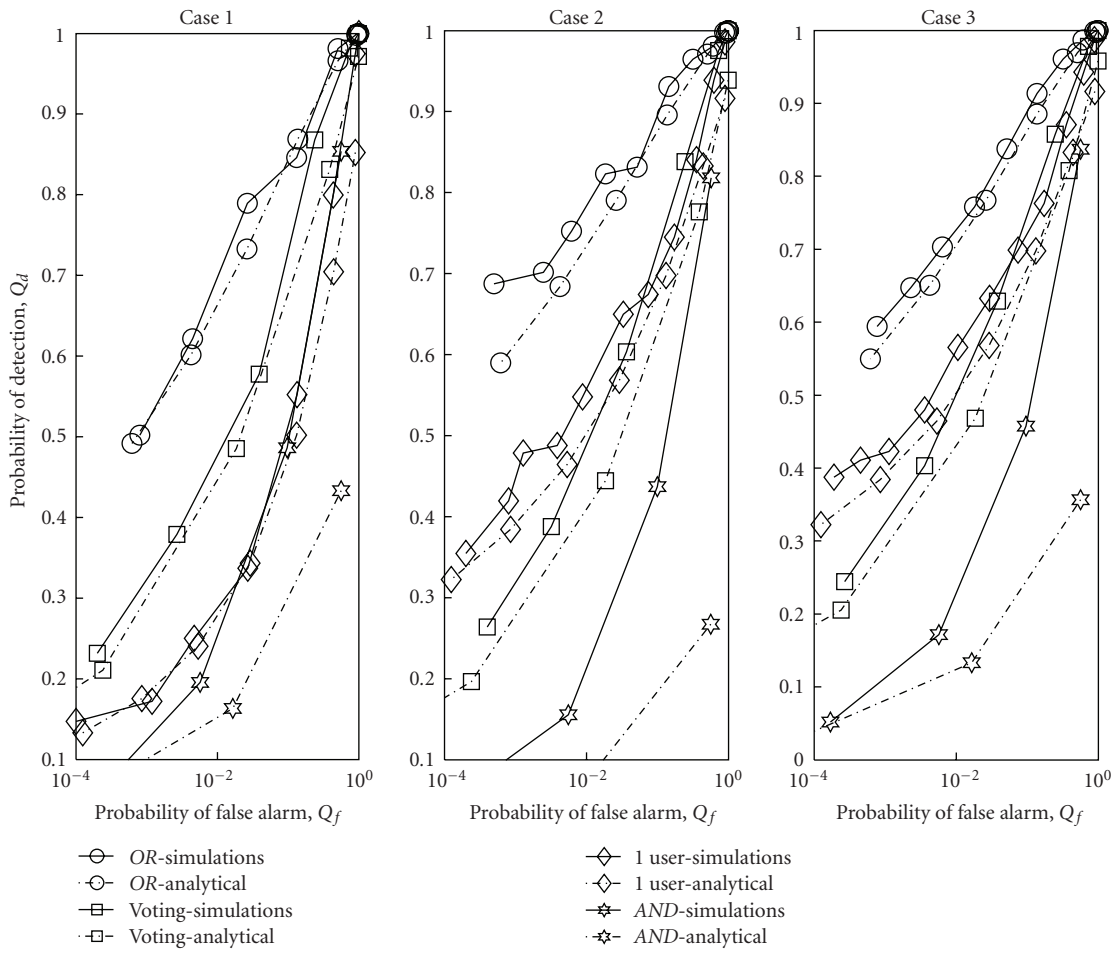


FIGURE 9: Receiver operating characteristics for 5 collaborating users in shadow fading ($\sigma_{\text{dB}} = 6$) in 3 cases: Case 1 (all users have similar SNR), Case 2 (half of the users have high SNR), and Case 3 (only one user has high SNR).

centre can be expressed as a function of the probability of detection (or false alarm) of each SU, obtaining the joint probability of M independent events as

$$\begin{aligned} Q_d &= R(\mathbf{D}) \prod_{S_0} \left(1 - P_d^i\right) \prod_{S_1} P_d^i, \\ Q_f &= R(\mathbf{D}) \prod_{S_0} \left(1 - P_f^i\right) \prod_{S_1} P_f^i. \end{aligned} \tag{23}$$

S represents the set of all secondary users with $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$ where $\mathcal{S}_0$ is the group of SUs that has decided that PU signal is absent while $\mathcal{S}_1$ is the group of SUs that has decided that PU signal is absent and $R(\mathbf{D})$ is the decision fusion rule at the fusion centre. Value of $R(\mathbf{D})$ depends on what type of fusion rule is used at the fusion centre. So for the given formulation, K-out-of-M rule can be formulated as

$$R(\mathbf{D}) = \begin{cases} 1 \text{ (PU present)} & \text{if } \sum_{i=1}^{M} D_i \geq K, \\ 0 \text{ (PU absent)} & \text{if } \sum_{i=1}^{M} D_i < K. \end{cases} \tag{24}$$

For the fusion rules considered in this section, K is given as

$$K = \begin{cases} 1, & \text{OR}, \\ M, & \text{AND}, \\ \left\lceil \dfrac{M}{2} \right\rceil, & \text{Voting}, \end{cases} \tag{25}$$

where $\lceil \cdot \rceil$ is the ceil function.

For *1-user* rule,

$$R(\mathbf{D}) = \begin{cases} 1, & \text{if } D_i = 1, \\ 0, & \text{otherwise}, \end{cases} \tag{26}$$

where the ith user is chosen as

$$i = \arg\max_j \left\{ \gamma_j \right\}. \tag{27}$$

5.1.2. Simulation Results. Figure 7 shows collaborative spectrum sensing performance when 5 SUs collaborate with each other and make collaborative decision; analytical results validating the simulation results are shown. When all users

have similar γ_i (Case 1) in AWGN channel, then the optimal decision fusion rule is *Voting* rule as evident from Figure 7. When half of the users have high γ_i and half of the users have low γ_i (Case 2), then the optimal decision rule is *OR*. Case 3 refers to the situation when only one user has higher γ_i than others; in this case the collaborative spectrum sensing works even worse than a single node. From these results it can be concluded that it is not necessary that collaboration of users always improves spectrum sensing performance and in order to derive an optimum performance individual γ_i should be considered. Without knowing γ_i the performance is always suboptimal; so in the proposed scenario it is suggested that users estimate local γ_i and send this information along with their 1-bit decision. Local SNR can be estimated by using a test statistics defined in (5) as

$$\hat{\gamma}_i = \frac{1}{Z}\sum_{i=1}^{Z} u_i - X, \tag{28}$$

where $\hat{\gamma}$ is estimated SNR, Z is the number of test statistics, and X is $\mathrm{E}(u_i \mid \mathcal{H}_0)$.

In Rayleigh fading and shadowing, collaborative spectrum sensing is an ideal solution because diversity gain achieved by collaboration effectively cancels the deleterious effects of fading. Figures 8 and 9 show detection performance under Rayleigh fading and shadowing with the three cases considered in this section. Value of dB-spread is assumed to be 6 dB for the shadowing while other parameters remain the same. As can be seen from these results, in all three cases spectrum sensing performance is superior if *OR* fusion rule is used at the fusion centre. So it can be concluded from simulation results that under Rayleigh fading and shadowing with i.i.d. measurements most optimal fusion rule is *OR* rule and collaboration of users is required. Further, with the increase of shadowing, sensing performance of two user collaboration with highest γ_i is better than collaboration of all users. It can be concluded that even in fading or shadowing it is important for the fusion centre to know the SNR values of the users to make a decision about which fusion rule gives better performance.

5.2. Spatially Correlated Shadowing. Up to this point, it is assumed that all collaborating cognitive users have identical and independent shadowing. However, usually there is a degree of spatial correlation associated with log-normal shadowing [52] and assumption of identically and independent (i.i.d.) shadowing is not always true. In this section, the impact of spatially correlated shadowing on decision fusion when users have different SNR is studied under different channel conditions. It is concluded that correlation has a direct impact on the optimum decision fusion rule at the fusion centre.

It is logical to think that spatially correlated shadowing would degrade the performance of CSS because such users are likely to experience similar observations thereby countering collaborative gains. In this paper correlated shadowing is modelled using the exponential correlation model [52]:

$$r(d) = e^{-ad}, \tag{29}$$

where $r(d)$ is the correlation matrix, d is the distance between two secondary users, and a is a constant depending on the environment. Based on measurements reported in [52], $a \approx 0.12$/m for urban environment and $a \approx 0.002$/m for suburban environment.

5.2.1. Mathematical Formulation. Assume that γ_i is the received SNR at the ith SU on a logarithmic scale. Hence under shadow fading γ_i has a Gaussian distribution with variance of σ^2_{dB} and a mean value of μ_γ (in dB). The value of μ_γ is determined by the distance dependent path loss. Under two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ the distribution of γ_i for M SUs under spatially correlated shadowing can be expressed as

$$\boldsymbol{\gamma}_{\mathrm{dB}} \sim \begin{cases} \mathcal{N}\left(0 \times \mathbf{u}_M, \sigma^2_{\mathrm{dB}}\boldsymbol{\Xi}\right), & \mathcal{H}_0, \\ \mathcal{N}\left(\boldsymbol{\mu}_\gamma, \sigma^2_{\mathrm{dB}}\boldsymbol{\Xi}\right), & \mathcal{H}_1, \end{cases} \tag{30}$$

where $\boldsymbol{\gamma}_{\mathrm{dB}} = [\gamma_1, \gamma_2, \ldots, \gamma_M]^T$, $\mathbf{u}_M$ is an $M \times 1$ vector of all ones, and $\boldsymbol{\Xi}$ is the normalised covariance matrix of $\boldsymbol{\gamma}_{\mathrm{dB}}$. Using the exponential correlation model defined in (29), the covariance matrix $\boldsymbol{\Xi}$ is an $M \times M$ matrix. Assuming that all SUs are uniformly distributed in a 1-dimensional plane within a total distance of κ, the elements of covariance matrix are given as

$$\boldsymbol{\Xi}_{i,j} = e^{(-a\kappa/(M-1))|i-j|}. \tag{31}$$

Hence, the covariance matrix $\boldsymbol{\Xi}$ can be expressed as

$$\boldsymbol{\Xi} = \begin{bmatrix} 1 & \mathcal{A} & \mathcal{B} & \cdots & e^{-a\kappa} \\ \mathcal{A} & 1 & \mathcal{A} & \cdots & e^{-a\kappa|M-2|/(M-1)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ e^{-a\kappa} & \mathcal{B} & e^{-3a\kappa/(M-1)} & \cdots & 1 \end{bmatrix}, \tag{32}$$

where $\mathcal{A}$ denotes $e^{-a\kappa/(M-1)}$ and $\mathcal{B}$ denotes $e^{-2a\kappa/(M-1)}$.

The probability density function of γ_i can be expressed for the M collaborative SUs having correlated shadow fading as

$$f\left(\boldsymbol{\gamma}_{\mathrm{dB}}\right) = \frac{1}{\sqrt{2\pi}\sigma^2_{\mathrm{dB}}}\boldsymbol{\Xi}^{-1}\exp\left\{-\frac{\left(\boldsymbol{\gamma}_{\mathrm{dB}} - \mu_\gamma\right)^2}{2\sigma^2_{\mathrm{dB}}}\boldsymbol{\Xi}^{-1}\right\}. \tag{33}$$

From (32) it is clear that $\boldsymbol{\Xi}$ is a diagonal constant matrix or Toeplitz matrix, and its inverse may be expressed as [53]

$$\boldsymbol{\Xi}^{-1} = \frac{1}{1 - e^{2a\kappa/(M-1)}}\begin{bmatrix} 1 & -\mathcal{A} & 0 & \cdots & 0 \\ -\mathcal{A} & 1 + e^{-2a\kappa/(M-1)} & -\mathcal{A} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & -\mathcal{A} & 1 \end{bmatrix}, \tag{34}$$

where $\mathcal{A}$ denotes $e^{-a\kappa/(M-1)}$.

5.2.2. Simulation Results. It is shown in this section that spatial correlation among users directly impacts the decision fusion at the fusion centre. Figures 10, 11, and 12 show ROC

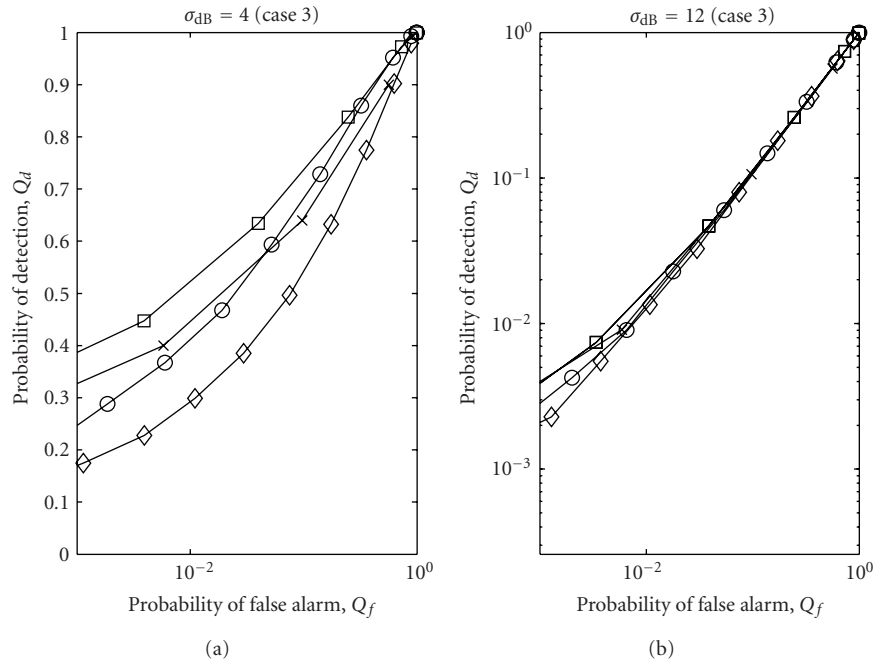


Figure 10: Receiver operating characteristics for 5 collaborating users in spatially correlated shadowing fading (Case 1: all users have similar SNR).

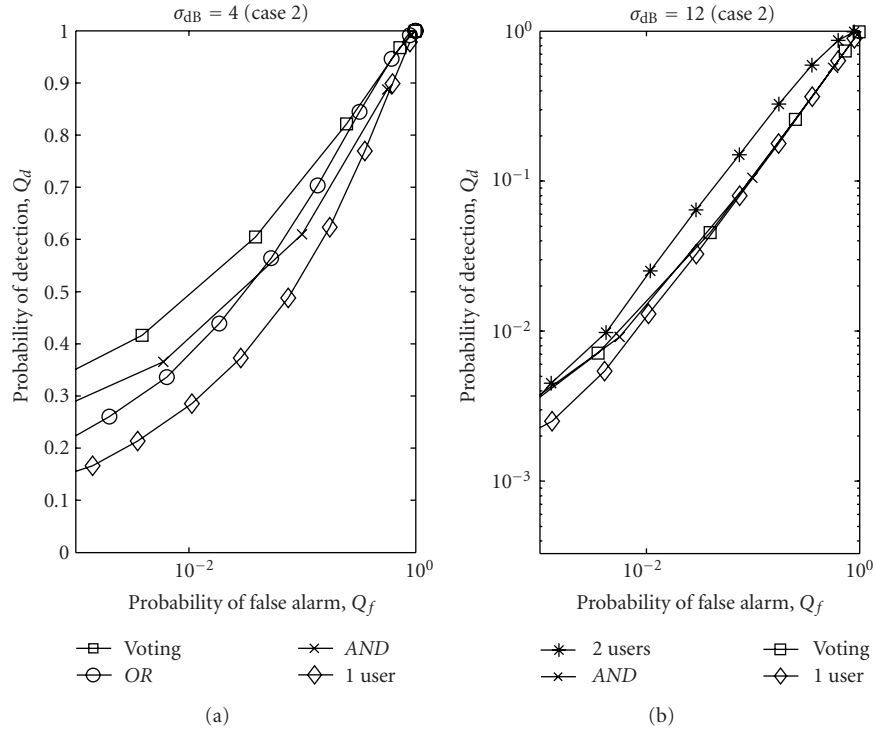


Figure 11: Receiver operating characteristics for 5 collaborating users in spatially correlated shadowing fading (Case 2: half of the users have high SNR).

curves of 5 collaborating users under spatially correlated shadowing with dB-spread of 4 dB and 12 dB for the three cases defined in Section 5.1. In case of correlated shadowing with lower values of σ_{dB}, the *Voting* fusion rule outperforms *OR* fusion rule and performance of *AND* fusion rule is better than *OR*. This is due to the fact that all secondary users are close to each other and have similar values of γ_i; hence user observations are similar to each other. However, sensing performance in heavily shadowed environment (e.g., when $\sigma_{dB} = 12$) for all fusion schemes is almost similar in all

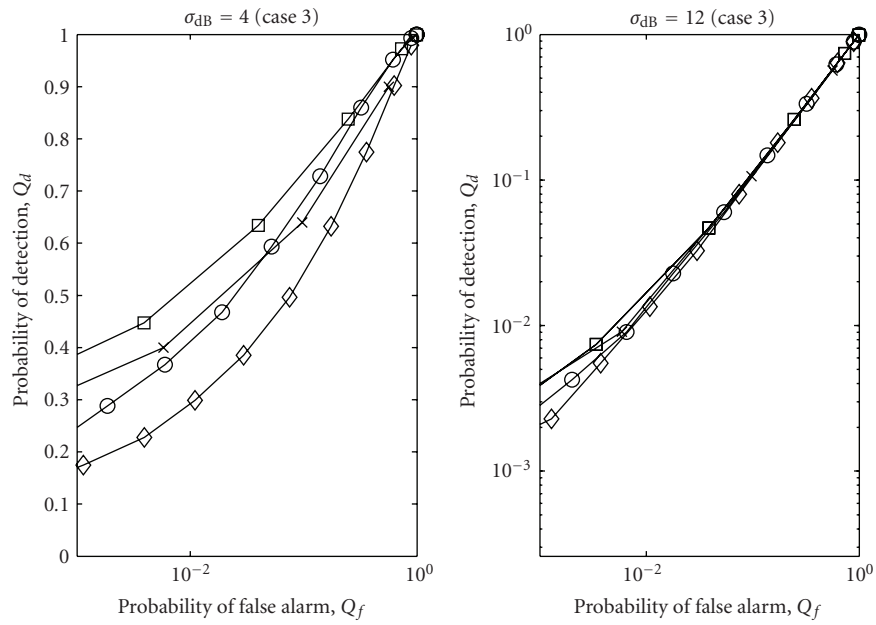


Figure 12: Receiver operating characteristics for 5 collaborating users in spatially correlated shadowing fading (Case 3: only one user has high SNR).

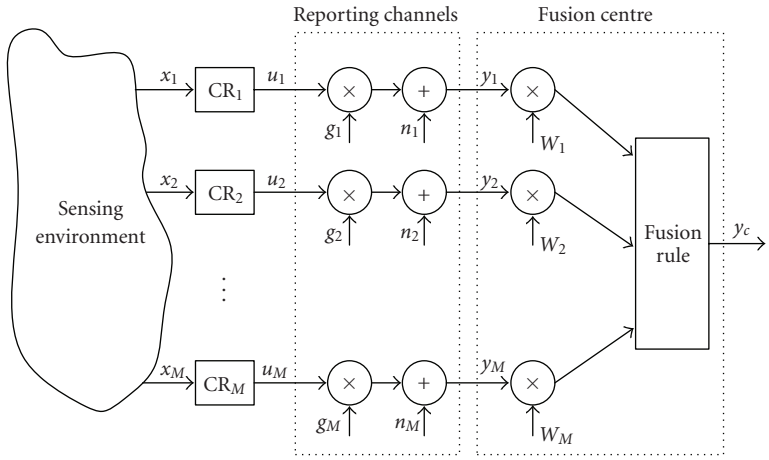


Figure 13: Schematic diagram of weighted collaboration at fusion centre for soft decision combining with imperfect reporting channels.

three cases. Hence, it can be seen from Figures 10, 11, and 12 that for optimal decision fusion at the fusion centre, it is important to consider the effects and degree of correlation among users.

6. Optimised User Collaboration Scheme for SDC

In this section, goal is to optimise CSS when collaborating SUs send their soft decisions to the fusion centre by maximising the global probability of detection (or alternatively minimising global probability of miss detection) for a given value of probability of false alarm and channel conditions. Referring to Section 4.2 for the framework of soft decision combing at the fusion centre, global probability of detection can be written in terms of global probability of false alarm (using (22)):

$$Q_d = Q\left(\frac{\sqrt{\mathbf{w}^T\left[2N\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}\,Q^{-1}\left(Q_f\right)}{\sqrt{\mathbf{w}^T\left[2(N\mathbf{I}+2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}} + \frac{N\mathbf{g}^T\boldsymbol{\Sigma}\mathbf{w}-\mathbf{g}^T((N\mathbf{I}+\boldsymbol{\Gamma})\odot\boldsymbol{\sigma})\mathbf{w}}{\sqrt{\mathbf{w}^T\left[2(N\mathbf{I}+2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}}\right). \tag{35}$$

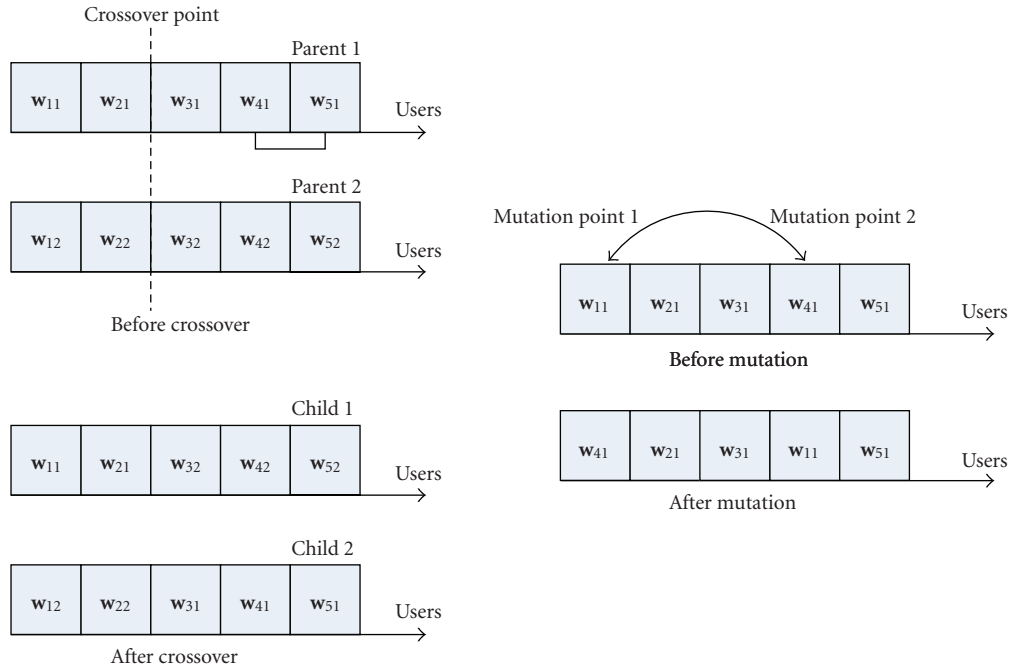


Figure 14: Crossover and mutation operations in genetic algorithm.

Maximising Q_d, as defined in (35), is equivalent to minimise $\varphi(\mathbf{w})$ as $Q(x)$ is a decreasing function of x, where $\varphi(\mathbf{w})$ is given by

$$\begin{aligned}\varphi(\mathbf{w}) &= \frac{\sqrt{\mathbf{w}^T\left[2N\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}\,Q^{-1}\left(Q_f\right)}{\sqrt{\mathbf{w}^T\left[2(N\mathbf{I}+2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}} \\ &\quad + \frac{N\mathbf{g}^T\boldsymbol{\Sigma}\mathbf{w}-\mathbf{g}^T((N\mathbf{I}+\boldsymbol{\Gamma})\odot\boldsymbol{\sigma})\mathbf{w}}{\sqrt{\mathbf{w}^T\left[2(N\mathbf{I}+2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}} \\ &= \frac{\sqrt{\mathbf{w}^T\left[2N\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}\,Q^{-1}\left(Q_f\right)-\boldsymbol{\Gamma}\mathbf{g}^T\boldsymbol{\Sigma}\mathbf{w}}{\sqrt{\mathbf{w}^T\left[2(N\mathbf{I}+2\boldsymbol{\Gamma})\mathbf{G}^2\boldsymbol{\Sigma}^2+\boldsymbol{\Delta}\right]\mathbf{w}}}.\end{aligned} \tag{36}$$

Similarly for fading channels, the average probability of detection can be obtained by averaging Q_d over fading statistics as described in Section 3.2. Now the optimisation problem can be formulated as

$$\begin{aligned}&\text{minimise} \quad \varphi(\mathbf{w}) \\ &\text{s.t.} \quad \sum_{i=1}^{M} w_i = 1, \quad w_i \geq 0, \ \forall i \in \{1,2,3,\ldots,M\}.\end{aligned} \tag{37}$$

6.1. GA-Based Weighted Collaborative Spectrum Sensing. This section describes the design of a GA-based weighted CSS framework for the case of SDC at the fusion centre. In this work, GA is used as a solution approach to minimise $\varphi(\mathbf{w})$ as defined in (36) for a given value of Q_f. The GA has been proposed as a computational analogy of adaptive systems by Holland [54]. They are modelled based on the principles of natural evolution and selection and is briefly described in this section. An initial population is first generated and then the fitness of each chromosome in the initial population is evaluated using a predefined fitness function. A loop is initiated to simulate the generations and in each generation, chromosomes are selected probabilistically according to their fitness. The genes of the selected individuals will mutate and crossover to produce offsprings to maintain the population size. The GA continues to iterate until the convergence is achieved or until it exceeds the maximum number of generations.

6.1.1. Seeding. The algorithm starts by randomly generating an initial population of possible solutions. Here, the initial population is the randomly generated values of weights satisfying the constraints as described in (37). Seeding is a process of setting the initial population to some initial configuration. If the initial population is seeded properly, the performance of GA can be greatly enhanced. Since GA works by probabilistically mutating and combining, the convergence of algorithm can be achieved quickly if the population is initially preset to a good solution.

6.1.2. Fitness Function. A fitness function plays a central role in GA. It evaluates fitness of each chromosome and forces the algorithm to search for optimal solutions and is the only link between actual problem and the GA. A fitness function ranks chromosomes in a given population; so individuals having better fitness values have higher chances of survival and reproduction in the next generation. In this paper,

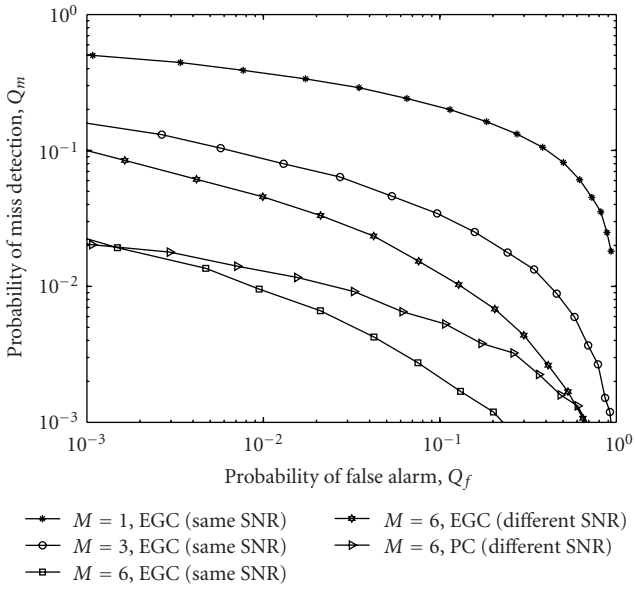


Figure 15: Receiver operating characteristics for collaborative spectrum sensing with perfect reporting channel, $N = 10$.

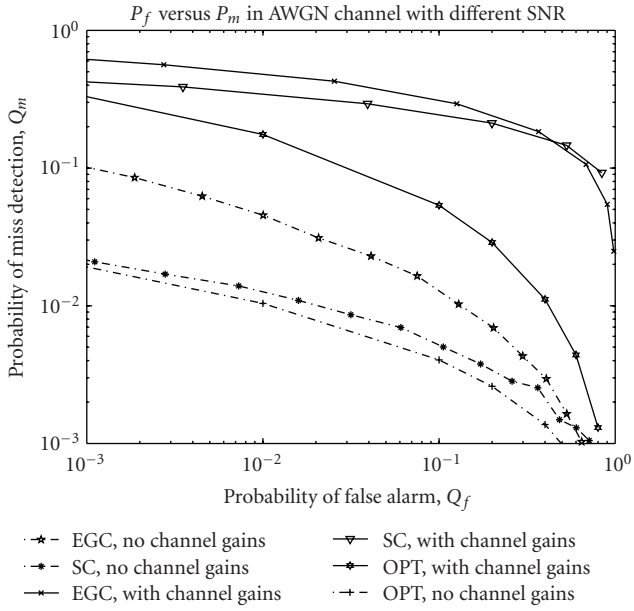


Figure 16: Receiver operating characteristics for collaborative spectrum sensing for 6 users in AWGN channel with im-perfect reporting channels, $N = 10$.

$\varphi(\mathbf{w})$ is used as a fitness function to evaluate the fitness of individuals. After calculating the fitness of each individual, all fitness values are scaled in a range that is suitable for the selection algorithm. The selection algorithm uses these scaled fitness values to choose the parents of the next generation. The range of scaled values affects the performance of GA, and in this paper the scaling method described by Goldberg [55] is used.

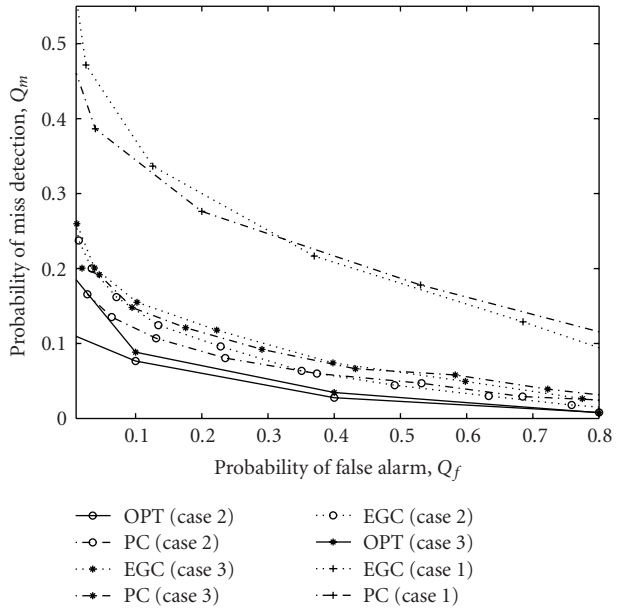


Figure 17: Receiver operating characteristics for collaborative spectrum sensing for 6 users in Rayleigh fading channel with im-perfect reporting channels, $N = 10$ (Case 1: All users have good reporting channel, Case 2: All users have bad reporting channel, and Case 3: Two of the users have good reporting channel).

6.1.3. Selection. Once the chromosomes in a given population have been evaluated according to their fitness values, the one with the better fitness will be selected, and the others will be eliminated. There are many different strategies available in the literature to implement selection algorithm [56]. The simplest and the most widely used selection scheme, the roulette wheel selection, is used in simulations [55].

6.1.4. Elitism. The number of chromosomes in a population with the best-scaled fitness value guaranteed to survive in the next generation and represented by Elitism is called Elite children. Proper value of Elite children is important in the fast convergence of GA.

6.1.5. Crossover and Mutation. Crossover process in GA combines two individuals (parents) and produces entirely new chromosomes (children). The main idea behind crossover operation is that the children may be better than both of the parents if they take the best attributes from each parent. Generally, crossover occurs during evolution according to a specified probability and is typically in the range of 80% to 90%. Although a number of crossover techniques are available in the literature, the simplest crossover technique, called single-point crossover [56], is used in this study.

Mutation is another genetic operation which alters one or more genes in a chromosome from its original state. This introduces new genetic material in the population. With the new gene, GA may be able to arrive at a better solution than previously possible. Mutation also occurs during the evolution process by some prespecified probability and this value is normally small as compared to the crossover

Table 1: GA parameter configuration.

Parameter	Value
Population Size	100
Number of Generations	40
Elitism	2%
Mutation Probability	2%
Crossover Probability	80%
Initialisation Method	Random
Crossover operation	Single point
Selection Method	Roulette wheel

probability. In the classical mutation process, one or more pairs of genes are selected randomly and swapped to produce new offsprings. Figure 14 illustrates the process of crossover and mutation for the case of 5 collaborating users.

6.1.6. Termination. Termination is the criteria by which GA decides whether to continue or stop searching for better solutions. There are many possibilities to terminate GA including generation number, evolution time, and population convergence, and so forth and in this work generation number criterion is used for terminating the GA. Based on a number of test experiments, the best suited GA parameter configuration was set up for the optimisation problem and parameters are listed in Table 1.

6.2. Numerical Result and Discussions. In this section, proposed GA-based weighted CSS scheme for SDC is simulated and compared with existing weighting schemes proposed in [42], that is, Equal Gain Combining (EGC) and Proportional Combining (PC). EGC is the weighting scheme in which all the collaborating SUs have equal weights and in PC the fusion centre assigns proportional weight to SUs according to their SNR values. Numerical results are obtained from simulations for use Case 2 over $1,000,000$ noise realisations for the given set of noise variances. Noise variance of all collaborating users for the primary channel (i.e., channel between primary transmitter and secondary users) is assumed to be $\sigma^2 = 1$ and noise variance of the reporting channels is assumed to be $\delta^2 = 1$ dB. Value of N is assumed to be 10 in all simulations.

Figure 15 shows the probability of miss detection Q_m against probability of false alarm Q_f with different number of collaborating users and their corresponding SNR values. A perfect reporting channel is assumed here and the channel between SUs and PU is considered to be AWGN channel. Figure 15 shows clearly that with an increase in the number of collaborating users sensing performance improves if all SUs have same SNR. However, when the cognitive users have different mean SNR values, then the sensing performance degrades with equal gain combining. Proportional weights assigned to different users according to their SNR values improve sensing performance as compared to equal gain combining approach. From the results it is concluded that users SNRs have a direct impact on the spectrum sensing performance.

Figure 16 plots the Q_m versus Q_f for the case when cognitive users have different SNRs and the reporting channel is not perfect; that is, practical AWGN channels exist between SUs and the fusion centre with different channel gains defined as $\mathbf{g} = [0.32, 0.2, 0.2, 0.1, 0.3, 0.15]^T$. The value of channel gain is dependent on the location of the fusion centre and the SU and is varying over time. It can be seen from Figure 16 that reporting channel gains degrade the performance of spectrum sensing. Without channel gains, PC performs better than EGC, but, in the presence of reporting channel, PC does not perform much better than EGC. This is mainly because of the fact that in the presence of imperfect reporting channel, optimum weights of cognitive users are not only dependant on SNR values but also depend on reporting channel conditions. Under such conditions an analytical expression for the probability of detection is derived and optimum weights are calculated using GA. The result shows that the proposed GA-based optimal weights, denoted as "OPT," yield superior spectrum sensing performance in both cases, that is, with and without reporting channel gain.

In order to evaluate the performance of proposed optimised collaborative spectrum sensing framework, performance of GA-based optimisation algorithm is tested in fading channel. Three different cases were considered: Case 1 refers to the case in which all the SUs have good reporting channel, Case 2 is the case in which all the collaborating cognitive users have bad reporting channel, while in Case 3 two of the collaborating users have strong channel, while others have bad reporting channel. As seen from Figure 17 spectrum sensing performance is the worst for Case 1 and the best for Case 3; however, in all of the three cases, the performance of the proposed optimised thresholds outperforms the other solutions.

7. Conclusions

Spectrum is a scarce resource and it has been a major focus of research over the last several decades. Cognitive radio technology, which is a one of the promising approaches to utilise radio spectrum efficiently, has become an attractive option. Deployment of cognitive radio networks mainly depends on the ability of cognitive devices to detect licensed or primary users accurately and hence minimise interference to the licensed users. Spectrum sensing has been identified as a key functionality of a cognitive radio. However, as observations of a single cognitive radio are not always trustworthy, so collaboration of cognitive users is normally required to improve licensed users detection performance. In this paper, optimisation algorithms for both hard decision and soft decision combining are presented for collaborative spectrum sensing. It is well known that fusion strategy at the fusion centre has direct impact on the overall performance of collaborative spectrum sensing. We consider optimisation of both hard and soft decision fusion and develop algorithms to optimise spectrum sensing performance. It is concluded

that in order to derive an optimum fusion rule, the fusion centre must know the collaborating users estimated SNR values, channel conditions, as well as their 1-bit decision for the case of hard decision fusion. We also proposed a genetic algorithm-based optimisation of weighted collaborative spectrum sensing in which weights are assigned to the information provided by the users to improve CSS in terms of ROC. The optimum weight vector is obtained by maximising the global probability of detection at the fusion centre. Simulation results show that the proposed strategies improve spectrum sensing performance in terms of global probability of miss detection. However, proposed schemes require knowledge about SNR of all users, channel conditions, reporting channel gains, and so forth which need larger bandwidths. Our future research will consider efficient protocols and techniques to optimise bandwidth utilisation for the cases presented in this paper.

Appendices

A. Derivation of Probability of False Alarm for Energy Detector

The probability distribution function of a chi-square random variable X with $2N$ degrees of freedom is given by

$$f_X(x) = \frac{x^{N-1}e^{-x/2}}{2^N\Gamma(N)}, \tag{A.1}$$

where $\Gamma(\cdot)$ is gamma function and is defined as

$$\Gamma(u) = \int_0^\infty a^{u-1}e^{-t}dt. \tag{A.2}$$

Now for a given threshold λ the probability of false alarm under hypothesis $\mathcal{H}_0$ (as defined in (1)) can be computed as

$$\begin{aligned} P_f &= \text{Prob}\{X > \lambda \mid \mathcal{H}_0\} \\ &= \int_\lambda^\infty f_X(x)dx \\ &= \int_\lambda^\infty \frac{x^{N-1}e^{-x/2}}{2^N\Gamma(N)}dx. \end{aligned} \tag{A.3}$$

Let $x = 2u$; so,

$$\begin{aligned} P_f &= \frac{1}{2^N\Gamma(N)}\int_{\lambda/2}^\infty 2^{N-1}u^{N-1}e^{-u}2\,du \\ &= \frac{1}{\Gamma(N)}\int_{\lambda/2}^\infty u^{N-1}e^{-u}du. \end{aligned} \tag{A.4}$$

From the definition of incomplete gamma function $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$,

$$P_f = \frac{\Gamma(N,\lambda/2)}{\Gamma(N)}. \tag{A.5}$$

B. Derivation of Probability of Detection for Energy Detector

Probability density function of noncentral chi-square random variable x with $2N$ degrees of freedom and noncentrality parameter of $2N\gamma$ is given by

$$f_X(x) = \frac{1}{2}\left(\frac{x}{2N\gamma}\right)^{(N-1)/2}\exp\left(-\frac{x+2N\gamma}{2}\right)I_{N-1}\left(\sqrt{2N\gamma x}\right). \tag{B.1}$$

So for the threshold λ, probability of detection, that is, probability that $X > \lambda$ under $\mathcal{H}_1$, is given as

$$\begin{aligned} P_d &= \text{Prob}\{X > \lambda \mid \mathcal{H}_1\} \\ &= \int_\lambda^\infty f_X(x)dx \\ &= \int_\lambda^\infty \frac{1}{2}\left(\frac{x}{2N\gamma}\right)^{(N-1)/2}\exp\left(-\frac{x+2N\gamma}{2}\right)I_{N-1}\left(\sqrt{2N\gamma x}\right)dx. \end{aligned} \tag{B.2}$$

Assume $x = z^2$; then,

$$\begin{aligned} P_d &= \int_{\sqrt{\lambda}}^\infty \frac{1}{2}\left(\frac{z^2}{2N\gamma}\right)^{(N-1)/2} \\ &\qquad \times \exp\left[-\frac{z^2+2N\gamma}{2}\right]I_{N-1}\left(z\sqrt{2N\gamma}\right)2z\,dz \\ &= \int_{\sqrt{\lambda}}^\infty \frac{1}{(2N\gamma)^{(N-1)/2}}z\cdot z^{N-1} \\ &\qquad \times \exp\left[-\frac{z^2+\left(\sqrt{2N\gamma}\right)^2}{2}\right]I_{N-1}\left(z\sqrt{2N\gamma}\right)dz. \end{aligned} \tag{B.3}$$

Using definition of generalised Marcum Q-function,

$$Q_m(\alpha,\beta) = \frac{1}{\alpha^{m-1}}\int_\beta^\alpha x^m \exp\left[-\frac{x^2+\alpha^2}{2}\right]I_{m-1}(\sqrt{\alpha x})\,dx. \tag{B.4}$$

P_d can be expressed in terms of generalised Marcum Q-function, with $m = N$, $x = z$, $\alpha = \sqrt{2N\gamma}$, and $\beta = \sqrt{\lambda}$, as

$$P_d = Q_N\left(\sqrt{2N\gamma}, \sqrt{\lambda}\right). \tag{B.5}$$

Acknowledgments

This work was performed in the project E3 which has received research funding from the EU FP7 framework. This paper reflects only the authors views and the community is not liable for any use that may be made of the information contained therein. The contributions of colleagues from the E3 consortium are hereby acknowledged.

References

[1] "International Telecommunication Union: Radio Regulations Revised," 1994.

[2] "Technology Research Programme: Research and Development at Ofcom 2004/05," October, p.37, 2005, http://www.ofcom.org.uk/research/technology/overview/techrandd200405/.

[3] "Federal Communications Commission: Spectrum policy task force report," November 2002, http://fjallfoss.fcc.gov/.

[4] A. Shukla, "Cognitive radio technology—a study for Ofcom," Tech. Rep. 830000143, QinetiQ Ltd, Hampshire, UK, 2006.

[5] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, "XG dynamic spectrum sharing field test results," in *Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 676–684, May 2007.

[6] Spectrum Management Advisory Group (SMAG), "Position Paper on UMTS," March 2009, http://www.ofcom.org.uk/static/archive/ra/smag/papers/umts.htm.

[7] P. Leaves, *Dynamic spectrum allocation between cellular and broadcast systems*, Ph.D. dissertation, University of Surrey, 2004.

[8] "FCC Auctions March 2008," July 2009, http://wireless.fcc.gov/.

[9] A. Ghasemi and E. S. Sousa, "Opportunistic spectrum access in fading channels through collaborative sensing," *IEEE Journal of Communications*, vol. 2, no. 2, pp. 71–81, 2007.

[10] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[11] H. Urkowitz, "Energy detection of unknown deterministic signals," *IEEE Proceedings*, vol. 55, no. 4, pp. 523–531, 1967.

[12] A. Sonnenschein and P. M. Fishman, "Radiometric detection of spread-spectrum signals in noise of uncertain power," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 654–660, 1992.

[13] Federal Communications Commission, "Notice of Proposed Rule Making, in the matter of unlicensed operaion in TV broadcast bands," May 2004, http://www.fcc.gov/.

[14] IEEE802.22, "Working Group on Wireless Regional Area Networks (WRAN)," Tech. Rep., IEEE, New York, NY, USA, 2009.

[15] Ofcom, "Digital Divident Review, A statement and our approach to awarding the digital divident," December 2007, http://www.ofcom.org.uk/consult/condocs/ddr/.

[16] DRIVE Project Web Site, March 2009, http://www.ist-drive.org/index2.html.

[17] Over-DRiVE Project Web Site, March 2009, http://www.ist-overdrive.org/.

[18] WINNER Project Web Site, March 2009, https://www.ist-winner.org.

[19] E2R II Project Web Site, March 2009, E2R II Project.

[20] ORACLE Project Web Site, March 2009, http://www.ist-oracle.org/.

[21] E3 Project Web Site, March 2009, https://ict-e3.eu/.

[22] Radio Access & Spectrum Cluster Web Site, March 2009, http://www.newcom-project.eu:8080/Plone/ras.

[23] DARPA neXt Generation (XG) Communications Working Group, March 2009, http://www.ir.bbn.com/projects/xmac/working-group/index.html.

[24] National Science Foundation, "Networking Technology and Systems (NeTS) Program Solicitation," March 2009, http://www.nsf.gov/pubs/2005/nsf05505/nsf05505.htm.

[25] P. Demestichas, A. Katidiotis, K. A. Tsagkaris, E. F. Adamopoulou, and K. P. Demestichas, "Enhancing channel estimation in cognitive radio systems by means of Bayesian networks," *Wireless Personal Communications*, vol. 49, no. 1, pp. 87–105, 2008.

[26] P. Demestichas, "Enhanced network selections in a cognitive wireless B3G world," *Annals of Telecommunications*, vol. 64, no. 7-8, pp. 483–501, 2009.

[27] Z. Xueqiang, C. U. I. Li, C. Juan, W. U. Qihui, and W. Jinlong, "Cooperative spectrum sensing in cognitive radio systems," in *Proceedings of the 1st International Congress on Image and Signal Processing (CISP '08)*, pp. 262–266, June 2008.

[28] W. Wang, W. Zou, Z. Zhou, H. Zhang, and Y. Ye, "Decision fusion of cooperative spectrum sensing for cognitive radio under bandwidth constraints," in *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology (ICCIT '08)*, pp. 733–736, December 2008.

[29] W. Zhang, R. K. Mallik, and K. Ben Letaief, "Cooperative spectrum sensing optimization in cognitive radio networks," in *Proceedings of IEEE International Conference on Communications (ICC '08)*, pp. 3411–3415, cn, June 2008.

[30] X. Chen, Z. s. Bie, and W. l. Wu, "Detection efficiency of cooperative spectrum sensing in cognitive radio network," *Journal of China Universities of Posts and Telecommunications*, vol. 15, no. 3, pp. 1–7, 2008.

[31] T. Aysal, S. Kandeepan, and R. Piesiewicz, "Cooperative spectrum sensing over imperfect channels," in *Proceedings of IEEE Globecom Workshops (GLOBECOM '08)*, pp. 1–5, December 2008.

[32] K. Arshad and K. Moessner, "Collaborative spectrum sensing for cognitive radio," in *Proceedings of IEEE International Conference on Communications Workshops (ICC '09)*, pp. 1–5, June 2009.

[33] E. Peh, Y. C. Liang, E. Peh, and Y. C. Liang, "Optimization for cooperative sensing in cognitive radio networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 27–32, April 2007.

[34] D. J. Kadhim, S. Gong, W. Liu, and W. Cheng, "Optimization of Cooperation sensing spectrum performance," in *Proceedings of WRI International Conference on Communications and Mobile Computing (CMC '09)*, pp. 78–82, cn, February 2009.

[35] E. Peh, Y.-C. Liang, and Y. L. Guan, "Optimization of cooperative sensing in cognitive radio networks: a sensing-throughput tradeoff view," in *Proceedings of IEEE International Conference on Communications*, June 2009.

[36] C. Sun, W. Zhang, and K. Letaief, "Cluster-based cooperative spectrum sensing in cognitive radio systems," in *Proceedings of IEEE International Conference on Communications (ICC '07)*, pp. 2511–2515, June 2007.

[37] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 1, pp. 98–101, 1986.

[38] F. F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, 2007.

[39] W. Wang, L. Zhang, W. Zou, and Z. Zhou, "On the distributed cooperative spectrum sensing for cognitive radio," in *Proceedings of International Symposium on Communications and Information Technologies (ISCIT '07)*, pp. 1496–1501, November 2007.

[40] E. Visotsky, S. Kuffher, and R. Peterson, "On collaborative detection of TV transmissions in support of dynamic spectrum sharing," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 338–345, December 2005.

[41] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proceedings of IEEE International Conference on Communications (ICC '06)*, pp. 1658–1663, August 2006.

[42] F. E. Visser, G. J. M. Janssen, and P. Pawełczak, "Multinode spectrum sensing based on energy detection for dynamic spectrum access," in *Proceedings of the 67th IEEE Vehicular Technology Conference (VTC '08)*, pp. 1394–1398, June 2008.

[43] Z. Quan, S. Cui, and A. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, 2008.

[44] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[45] A. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, NY, USA, 2005.

[46] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer, Secaucus, NJ, USA, 1996.

[47] R. F. Mills and G. E. Prescott, "A comparison of various radiometer detection models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 467–473, 1996.

[48] I. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40–48, 2008.

[49] T. Weiss, J. Hillenbrand, and F. Jondral, "A diversity approach for the detection of idle resources in spectrum pooling systems," in *Proceedings of the 48th International Scientific Colloquium*, pp. 37–38, September 2003.

[50] C. Da Silva, B. Choi, and K. Kim, "Distributed spectrum sensing for cognitive radio systems," in *Proceedings of the Information Theory and Applications Workshop (ITA '07)*, pp. 120–123, March 2007.

[51] A. K. Kattepur, A. T. Hoang, Y.-C. Liang, and M. J. Er, "Data and decision fusion for distributed spectrum sensing in cognitive radio networks," in *Proceedings of the 6th International Conference on Information, Communications and Signal Processing (ICICS '07)*, pp. 1–5, December 2007.

[52] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[53] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[54] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, Cambridge, Mass, USA, 1992.

[55] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Longman, Boston, Mass, USA, 1989.

[56] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Design*, Wiley, New York, NY, USA, 1997.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 160462, 8 pages
doi:10.1155/2010/160462

Research Article

Fast Detection Method in Cooperative Cognitive Radio Networks

Zhengyi Li, Lin Liu, and Chi Zhou

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60626, USA

Correspondence should be addressed to Chi Zhou, zhouc@ece.iit.edu

Received 1 December 2009; Revised 13 May 2010; Accepted 8 June 2010

Academic Editor: Hsiao Hwa Chen

Cognitive Radio (CR) technology improves the utilization of spectrum highly via opportunistic spectrum sharing, which requests fast detection as the spectrum utilization is dynamic. Taking into consideration the characteristic of wireless channels, we propose a fast detection scheme for a cooperative cognitive radio network, which consists of multiple CRs and a central control office. Specifically, each CR makes individual detection decision using the sequential probability ratio test combined with Neyman Pearson detection with respect to a specific observation window length. The proposed method upper bounds the detection delay. In addition, a weighted K out of N fusion rule is also proposed for the central control office to reach fast global decision based on the information collected from CRs, with more weights assigned for CRs with good channel conditions. Simulation results show that the proposed scheme can achieve fast detection while maintaining the detection accuracy.

1. Introduction

In the traditional management of licensed spectrum, users usually pay and have the exclusive access of spectrum with a certain level of Quality of Service (QoS) guarantee. On one hand, the spectrum is getting more and more crowded as the number of wireless devices increases drastically. However, on the other hand, the utilization of spectrum at any given time is low. Figure 1 shows a measurement of 30M–3GHz spectrum utilization. We can see that a lot of spectrum bands are vacant. Therefore, it would be efficient to allow unlicensed users to share spectrum with licensed users by using a vacant frequency band.

Cognitive Radio technology is developed to utilize these white spaces intelligently [1, 2]. FCC Spectrum Policy Task Force published a new spectrum management policy, open access or license exempted model, in 2002, to allow unlicensed user to use the opportunistic spectrum. As the transition from analog to digital television is complete, there are vacant channels (white spaces) in every media market [3]. Accordingly, the FCC announced a Notice of Proposed Rule Making (NPRM) on 13 May 2004, which proposed "to allow unlicensed radio transmitters to operate in the broadcast TV spectrum at locations where that spectrum is not being used". Seen as the secondary user, the cognitive radio (CR) must avoid interfering with primary user (PU), that is, licensed user, while sharing the licensed band with the PU. Therefore, cognitive radio needs to sense the spectrum to detect the existence of PU, identify the white spaces of spectrum, and adapt its transmission to one of the white spaces to avoid interfering with PU.

Detecting the vacant bands of the spectrum is the very first step but very crucial in Cognitive Radio technology. There are three major digital signal processing techniques that could be used to detect the existence of PU: matched filtering, energy detection, and cyclostationary feature detection [4, 5]. Among those, energy detector has been used widely due to its simplicity and easy implementation [6]. As a radio device, a single CR may suffer severe shadowing or multipath fading with respect to primary transmitter so that it cannot detect the existence of PU even in its vicinities. In addition, there exists a hidden-node problem, in which a CR may be too far from the PU to detect the existence, but close to the primary receiver to interfere with the reception if transmited. Cooperative sensing provides a solution to the challenges mentioned above [7, 8]. In cooperative sensing, multiple cognitive radios cooperate to reach an optimal global decision by exchanging and combining individual local sensing results. Allowing multiple CRs to cooperate, cooperative sensing can increase the detection probability, reduce the detection time, and achieve the diversity gain [9–18].

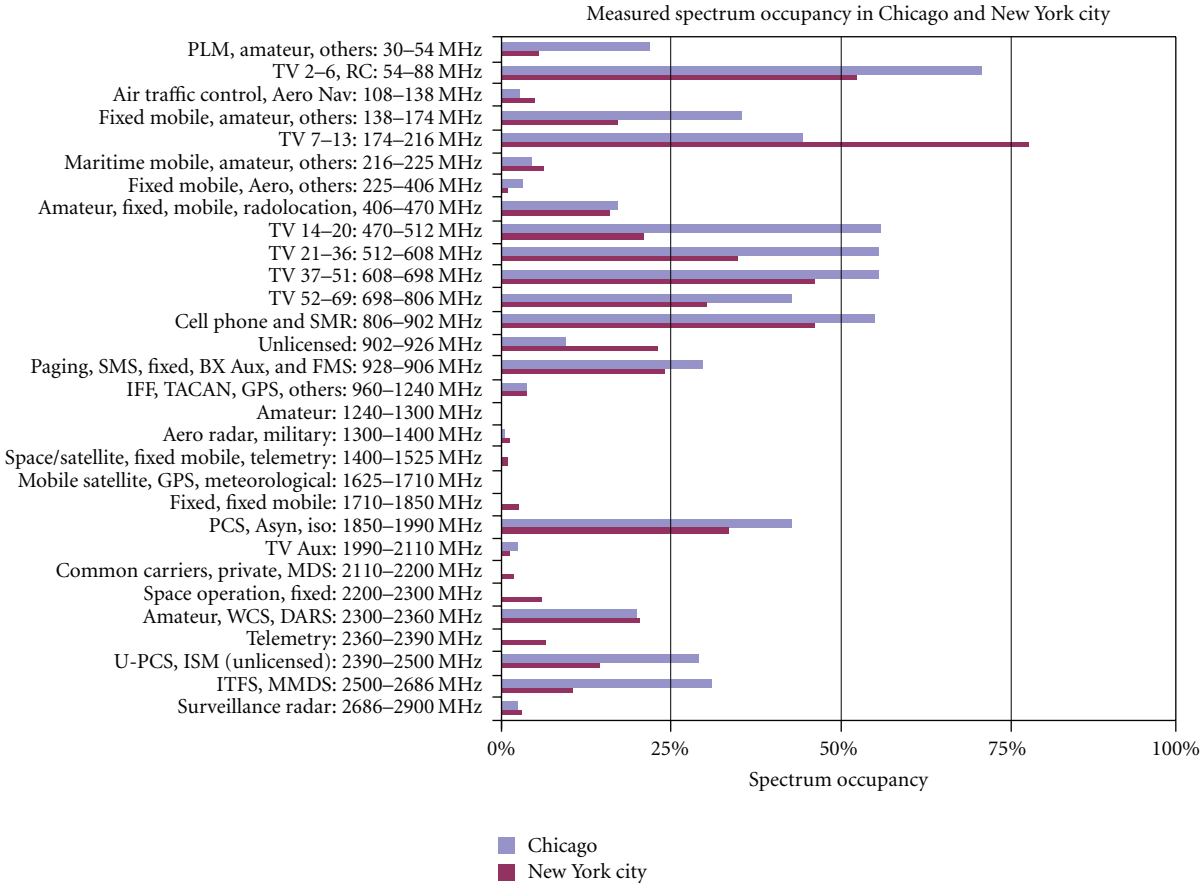


Figure 1: A measurement of 30M–3GHz spectrum utilization.

Due to the fading and noisy wireless channel, a large number of samples are needed for accurate detection. However, the spectrum utilization is dynamic, which requests fast detection to enable opportunistic sharing. In this paper, we propose a fast detection scheme in a cooperative cognitive radio network, which consists N CRs and a central control office. Each CR makes individual detection decision and then forwards its decision and the average signal to noise radio (SNR) to the central control office, which will make a global detection decision based on the collected data from CRs in the network. Then the central control office broadcasts the global detection decision to all the CRs. The proposed scheme consists of two folds: the first is to propose Sequential Probability Ratio Test (SPRT) method with a truncated window to upper bound the detection time at individual CR, while satisfying the detection accuracy requirements; the second is to propose a weighted K out of N fusion rule, which assigns more weights for CRs with good channel conditions, at the central control office to speed up the global decision making by using less number of individual decisions. Simulation results show that the proposed scheme can achieve fast detection while maintaining the detection accuracy.

The remaining contents are organized as follows. In Section 2, we discuss the system model. Section 3 presents the proposed fast detection scheme. In Section 4, simulation results are presented. In the end, we give the conclusion.

2. System Model

We consider a cognitive network, which consists of N CRs and a central control office. Each CR is equipped with an energy detector to individually detect the existence of PU by measuring the received SNR. Once the detection decision is reached by a CR, the CR transmits its decision along with the average received SNR to the central control office for global decision making. Serving as a fusion center, the central control office applies some fusion rule to its collected data and reach the global detection decision. Then the central control office broadcasts the global detection decision to all the CRs. Widely adopted Ad hoc On-demand Distance Vector (AODV) routing protocol [19] is used over a default clear channel for information exchange between the CRs and the central control office. The default channel may be selected among several predetermined channels.

In this paper, we use the log-normal shadowing path loss model:

$$P_r(d) = P_0(d_0) + 10 \cdot n \cdot \mathrm{Log}(d_0 \setminus d) + X(0, \delta), \quad (1)$$

where $P_r(d)$ is the received signal power at distance d, $P_0(d_0)$ is the received power at the reference point d_0, n is the path loss exponent, and $X(0, \delta)$ is normal shadowing random variable with zero mean and δ standard variance in dB.

We summarize the major notations which will be used in the paper in Table 1.

3. Fast Detection Scheme

In this section, a fast detection scheme is proposed and discussed in details. SPRT with truncated window is proposed for individual detection, followed by a weighted K out of N fusion rule for the central control office to reach quick global decision.

3.1. Individual Detection. Fast and accurate individual detection is a must. To achieve desired detection accuracy, multiple samples need to be taken due to the time-varying wireless links. One approach is to take a certain amount of samples and then make a one-time decision, such as Neyman Pearson method [20]. Another approach is sequential detection, that is, the detection decision criterion will be checked whenever one new sample is taken, such as Sequential Probability Ratio Test (SPRT) [21]. Neyman Pearson method has a fixed detection delay, while SPRT usually takes less-detection time on average but may take long delay though with small probability. We propose to combine these two approaches together to take advantages of the two. Specifically, we propose to impose a truncated window to SPRT so that the detection delay is bounded. When the number of the samples is less than the window size, original SPRT is used to do the sequential detection. If the sequential detection cannot reach decision when the window size is reached, Neyman Pearson method will be used to make the final decision. The proposed SPRT with truncated window achieves smaller detection delay compared with SPRT and Neyman Pearson method.

3.1.1. Sequential Probability Ratio Test. We define two hypotheses, specified as follows:

$$\begin{aligned} &H_0\text{: the primary user does not exist,}\\ &H_1\text{: the primary user does exist.} \end{aligned} \tag{2}$$

When a CR observes a new sample from energy detector, it will compute the cumulative sum of the log-likelihood ratio. We assume all samples are i.i.d. Let y_i be the received power from the ith observed sample, and then the log-likelihood ratio for the sample is

$$l(y_i) = \ln \frac{pdf(y_i \mid H_1)}{pdf(y_i \mid H_0)}. \tag{3}$$

When H_0 is true, that is, the PU does not exist, y_i is just the noise power. When H_1 is true, that is, the PU does exist, y_i is the received signal power plus the noise power. Let noise(i) be the AWGN noise for the ith sample with zero mean and variance δ_n^2. Then y_i is normal distributed

$$\begin{aligned} &\text{If } H_0 \text{ is true: } y_i = \text{noise}(i) \sim \text{Normal}(0, \delta_n^2),\\ &\text{If } H_1 \text{ is true: } y_i = u + \text{noise}(i) \sim \text{Normal}(u, \delta_n^2), \end{aligned} \tag{4}$$

where u is the signal power. Therefore, the log-likelihood ratio for the sample is

$$l(y_i) = \frac{uy_i}{\delta_n^2} - \frac{u}{2\delta_n^2}. \tag{5}$$

Table 1: Notation table.

λ_0,λ_1:	two stopping bounds in SPRT
$P_{m,\text{sprt}}$:	miss-detection probability in SPRT
$P_{d,\text{sprt}}$:	detection probability in SPRT
$P_{f,\text{sprt}}$:	false alarm probability in SPRT
$P_{d,\text{NP}}$:	detection probability in NP
$P_{f,\text{NP}}$:	false alarm probability in NP
P_{dd}:	detection probability in SPRT-TW
P_{ff}:	false alarm probability in SPRT-TW
$P(H_0)$:	statistic probability of H_0
$P(H_1)$:	statistic probability of H_1
w:	window size
I_i:	individual decision
α_i:	assigned weight value
SNR_i:	received SNR of the ith CR
d_0:	reference distance
n:	path loss exponent
$P_0(d_0)$:	received power at reference distance
P_{rc}:	confidence probability

The cumulative sum of the log-likelihood ratio can be written in the sequential way as

$$L(Y_i) = L(Y_{i-1}) + l(y_i), \tag{6}$$

where

$$L(Y_0) = 0. \tag{7}$$

It can be also written as

$$L(Y_i) = \sum_{k=1}^{i} l(y_k). \tag{8}$$

According to (5), we have

$$L(Y_i) = \sum_{k=1}^{i} l(y_k) = \frac{u}{\delta_n^2} \sum_{k=1}^{i} y_k - \frac{iu}{2\delta_n^2}. \tag{9}$$

The cumulative sum of the log-likelihood ratio will be compared with two stopping bounds, λ_0 and λ_1, to make decision. When the cumulative sum $L(Y_i)$ is larger than λ_1, we accept H_1 hypothesis and the detection process stops. If the cumulative sum $L(Y_i)$ is less than λ_0, we accept H_0 hypothesis and the detection process also stops. However, when $L(Y_i)$ lies between these two bounds, a new sample will be taken and the cumulative sum will be updated and compared with the bounds. The sequential detection process continues until it stops.

These two stopping bounds are set to satisfy the required miss-detection probability $P_{m,\text{sprt}}$ and false alarm probability $P_{f,\text{sprt}}$. They can be approximated as [21]

$$\begin{aligned} \lambda_0 &\approx \ln \frac{P_{m,\text{sprt}}}{1 - P_{f,\text{sprt}}},\\ \lambda_1 &\approx \ln \frac{1 - P_{m,\text{sprt}}}{P_{f,\text{sprt}}}. \end{aligned} \tag{10}$$

Let T be the detection time. We could obtain expected detection time:

$$E(T \mid H_1) = \frac{\left(1 - P_{m,\text{sprt}}\right)\lambda_1 + P_{m,\text{sprt}}\lambda_0}{E(l(y_i) \mid H_1)},$$
$$E(T \mid H_0) = \frac{\left(1 - P_{f,\text{sprt}}\right)\lambda_0 + P_{f,\text{sprt}}\lambda_1}{E(l(y_i) \mid H_0)}. \tag{11}$$

3.1.2. Proposed SPRT with Truncated Window. The sequential detection process is random and may take a very long time before it stops. In order to put an upper bound on the detection time, we impose a truncated window with size w to SPRT. If SPRT cannot stop within w samples, instead of taking more samples, we apply Neyman Pearson (NP) method to reach immediate decision while achieving certain false alarm probability $P_{f,\text{NP}}$ and detection probability $P_{d,\text{NP}}$ with w samples.

The proposed SPRT with truncated window (SPRT-TW) scheme is summarized as follows:

$$\begin{aligned} &L(Y_i) \geq \lambda_1\text{: Accept } H_1, \\ &L(Y_i) \leq \lambda_0\text{: Accept } H_0, \\ &\lambda_0 < L(Y_i) < \lambda_1 \text{ and } i < w\text{: Continue sampling}, \\ &\lambda_0 < L(Y_i) < \lambda_1 \text{ and } i = w\text{: Apply NP method.} \end{aligned} \tag{12}$$

Therefore, the individual detection probability P_{dd} and the individual false alarm probability P_{ff} for the proposed SPRT-TW could be written as, according to Bayes' Rule,

$$\begin{aligned} P_{dd} &= P_{d,\text{NP}}(w) \cdot P(T > w) + P_{d,\text{sprt}} \cdot P(T \leq w), \\ P_{ff} &= P_{f,\text{NP}}(w) \cdot P(T > w) + P_{f,\text{sprt}} \cdot P(T \leq w), \end{aligned} \tag{13}$$

where $P(T > w)$ is the probability that the CR does not reach a decision within window size w samples and $P(T \leq w)$ is the probability that the CR reaches a decision within window size w. According to the rule of total probability, the two probabilities can be expressed as

$$\begin{aligned} P(T > w) &= P(T > w \mid H_1) \cdot P(H_1) \\ &\quad + P(T > wH_0) \cdot P(H_0), \\ P(T \leq w) &= P(T \leq w \mid H_1) \cdot P(H_1) \\ &\quad + P(T \leq w \mid H_0) \cdot P(H_0) \end{aligned} \tag{14}$$

where $P(H_0)$ and $P(H_1)$ are statistical probabilities for the two hypothesis and

$$\begin{aligned} P(T > w \mid H_1) &= \prod_{i=1}^{w} P(\lambda_0 < L(Y_i) \leq \lambda_1 \mid H_1), \\ P(T > w \mid H_0) &= \prod_{i=1}^{w} P(\lambda_0 < L(Y_i) \leq \lambda_1 \mid H_0), \\ P(T \leq w \mid H_1) &= 1 - P(T > w \mid H_1), \\ P(T \leq w \mid H_0) &= 1 - P(T > w \mid H_0). \end{aligned} \tag{15}$$

From (9), $L(Y_i)$ is the sum of normally distributed random variables. Therefore, $L(Y_i)$ follows normal distribution

$$\begin{aligned} &\text{If } H_0 \text{ is true: } L(Y_i) \sim \text{Normal}(-b, a^2 i\delta_n^2), \\ &\text{If } H_1 \text{ is true: } L(Y_i) \sim \text{Normal}(aui - b, a^2 i\delta_n^2), \end{aligned} \tag{16}$$

where

$$a = \frac{u}{\delta_n^2}, \qquad b = \frac{iu^2}{2\delta_n^2}. \tag{17}$$

The expected detection delay for SPRT-TW could be obtained:

$$\begin{aligned} E(T \mid H_1) &= 1 \cdot P(T \leq 1 \mid H_1) \\ &\quad + \sum_{i=2}^{w} i \cdot P(T > i - 1 \mid H_1) P(T = i \mid H_1) \\ &\quad + w \cdot P(T > w \mid H_1) P_{d,\text{NP}}(w), \\ E(T \mid H_0) &= 1 \cdot P(T \leq 1 \mid H_0) \\ &\quad + \sum_{i=2}^{w} i \cdot P(T > i - 1 \mid H_0) P(T = i \mid H_0) \\ &\quad + w \cdot P(T > w \mid H_0) P_{f,\text{NP}}(w), \end{aligned} \tag{18}$$

where

$$\begin{aligned} P(T > w \mid H_1) &= 1 - P(\lambda_0 < L(Y_w) \leq \lambda_1 \mid H_1), \\ P(T > w \mid H_0) &= 1 - P(\lambda_0 < L(Y_w) \leq \lambda_1 \mid H_0). \end{aligned} \tag{19}$$

Based on (13)–(17), the expected delay could be obtained easily.

3.2. Weighted K out of N Fusion Rule. Data fusion is a technique used to efficiently combine the data for decision making. Due to its simplicity and effectiveness, K out of N fusion rule has been widely used in many applications including cognitive radio [7, 22]. We could also apply the K out of N fusion rule in the central control office to reach the global detection decision. Similar to [23], the global decision rule could be specified as

$$\begin{aligned} &\sum_N I_i \geq K\text{: Accept } H_1, \\ &\sum_N I_i < K\text{: Accept } H_0, \end{aligned} \tag{20}$$

where I_i is the indicator of individual detection decision for CR_i. $I_i = 1$ if CR_i accepts H_1 and $I_i = 0$ if CR_i accepts H_0.

The above K out of N fusion detection rule implies that each data has the same credibility as others by simply adding the individual detection decisions together. However, this is not true in wireless communication systems. For example, suppose two CRs correctly detect the existence of the PU with

one CR located very close to the PU and the other located far away from the PU. The nearby CR receives strong signal and quickly detects the PU, while the far-away CR receives very weak signal and takes much longer time to reach the decision. Obviously, the detection decision from the nearby CR is more reliable, which is not taken into account in the original K out of N fusion rule. Therefore, we propose a weighted K out of N fusion rule by assigning bigger weight to the CR with good signal reception (i.e., good channel condition). Then the global decision rule is specified as

$$\sum_{N} \alpha_i I_i \geq K\text{: Accept } H_1,$$
$$\sum_{N} \alpha_i I_i < K\text{: Accept } H_0, \tag{21}$$

where α_i is the weight for individual decision of CR_i. There are many ways to design the weight α_i to reflect the credibility of individual decision. In this paper, as an example, we design the weight as a linear function of received SNR

$$\alpha_i = A \cdot \text{SNR}_i + B, \tag{22}$$

where A and B are some constants.

In the weighted K out of N fusion rule, the individual detection decisions under good channel conditions are given more weights in the global decision making. Therefore, the global decision making requests a small number of CRs if those CRs have good channel condition or a large number of CRs if they have bad channel condition. Since those CRs with good channel condition also have smaller detection time using SPRT-TW and consequently their decisions arrive at the central control office faster, the global detection time can be reduced when most CRs have good channel condition.

4. Simulation Results

In this section, we first consider the individual detection performance for each CR, and then evaluate the weighted K out of N fusion rule for global detection. For the log-normal shadowing path loss model, the shadowing random variable $X(0, \delta)$, adds the randomness to the results, which complicates the illustration and insight discussion. Therefore, in the simulation for the individual detection, we first consider the log-distance path loss model without fading and then generalize it to slow fading scenario. The simulation results in the no fading scenario help understand the whole innovative fast detection scheme. Throughout simulation, we set the following parameters: the noise follows normal distribution with zero mean and the noise power set as -120 dBm; the path loss exponent $n = 4$; $P_0(d_0)$ is set as 20 dBm; reference distance is set $d_0 = 1$ m; the statistic probabilities $P(H_0) = P(H_1) = 0.5$.

4.1. Individual Detection Scenario 1: No Fading. With no fading, the path loss model is simplified to the log-distance path loss model and the average received signal power is

$$u = P_r(d) = P_0(d_0) + 10 \cdot n \cdot \text{Log}(d_0 \setminus d). \tag{23}$$

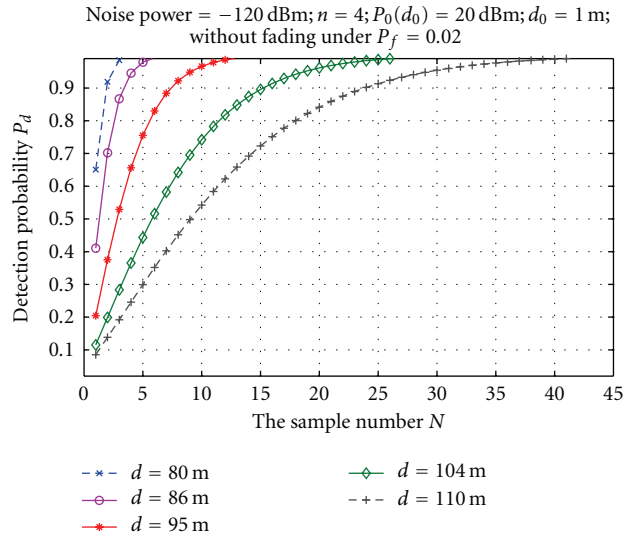


Figure 2: Traditional NP method: under fixed $P_{f,\text{NP}} = 0.02$.

CRs with longer distance to the PU receive weaker signal power, therefore, the distance can be used to represent the received signal strength or the received SNR.

Note that the detection performance depends on the window size w and received signal power u. Intuitively, the larger the window size is, the better the detection performs; the stronger the received signal power u is, the better the detection performs. We first examine the performance of traditional Neyman Pearson (NP) method when varying the number of samples (i.e., the window size w in SPRT-TW). We fix the false alarm probability for NP method as 0.05. Figure 2 shows how the detection probability of NP method varies with the number of samples for CRs at different distance from the PU. It is shown that the detection probability increases as the number of samples increases and the CR with smaller distance (i.e., stronger signal) achieves higher detection probability for any given number of samples. As shown in Figure 2, the traditional NP method needs up to 42 samples to reach 0.99 detection probability for CRs at the distance 110 m.

We set the two stopping bounds of SPRT based on the miss-detection probability $P_{m,\text{sprt}} = 0.01$ and false alarm probability $P_{f,\text{sprt}} = 0.02$ and simulate the proposed SPRT-TW. Figure 3 shows how the detection probability P_{dd} of SPRT-TW varies with the window size for CRs with different distance. Figure 3 has the same trend as Figure 2. Shown in Figure 3, the CR at distance 110 m (received SNR is 58.343 dB) takes most window size $w = 18$ to meet 0.99 detection probability while the CR located at $d = 80$ m (received SNR is 63.8764 dB) only needs 2 samples on average to reach the same detection probability. Compared with the NP method, SPRT-TW takes less samples to reach the same detection probability (e.g., 42 samples for NP method and 18 samples for SPRT-TW at the same CR at $d = 110$ m).

Figure 4 shows the corresponding false alarm probability P_{ff} of SPRT-TW as the window size varies. It is shown that

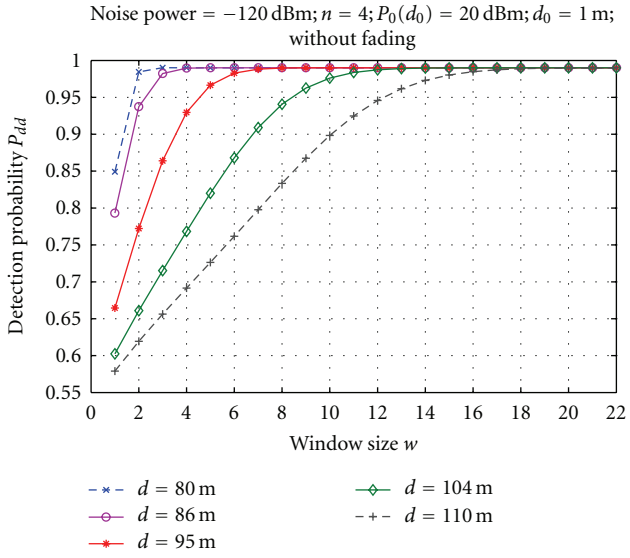


Figure 3: Detection probability P_{dd} versus Window size.

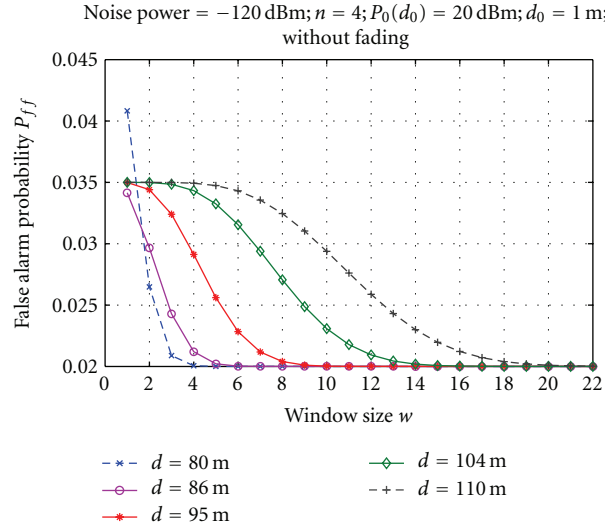


Figure 4: False alarm probability P_{ff} versus Window size

P_{ff} decreases with the window size and the CR at the smaller distance achieves smaller P_{ff} for any given window size. Shown in Figure 4, the CR at $d = 110$ m needs window size $w = 20$ to reach 0.02 false alarm probability, which takes only 4 window size for the CR at $d = 80$ m.

From Figures 3 and 4, we see that the window size needs to be carefully selected to meet the desired detection performance. In addition, the selection of window size also depends on the channel condition (i.e., received SNR). In order to minimize the detection delay, we need to select the minimum window size w that can meet the performance requirements.

We further compare the detection delay among NP method, SPRT, and our proposed SPRT-TW. Figure 5 shows that the NP method needs the longest delay and our proposed SPRT-TW has smallest delay. The simulation results validate that our SPRT-TW is indeed a fast detection scheme compared to NP method and original SPRT.

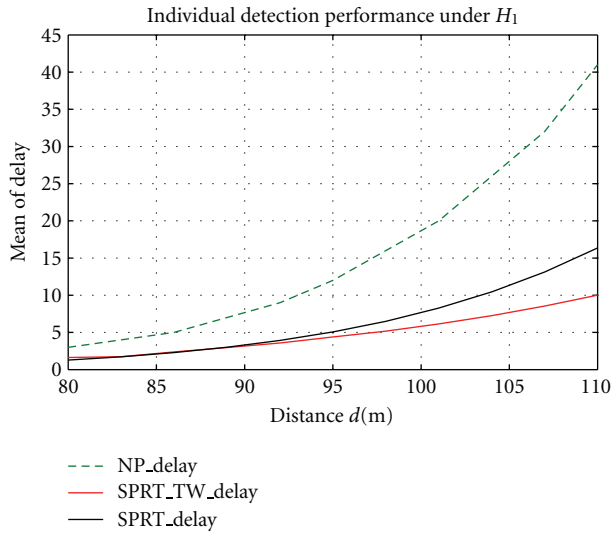


Figure 5: Delay performance Comparison.

4.2. Individual Detection Scenario 2: Slow Fading. In this section, we consider the log-normal shadowing path loss model, shown in (1). Because of the random fading factor $X \sim N(0, \delta(\text{dB}))$, we define a confidence probability Prc,

$$\text{Prc} = P(P_r(d) > \varepsilon), \tag{24}$$

where the $P_r(d)$ is received power and ε is the power threshold. Prc describes how much confidence we have when the receive power is stronger than the power threshold. In the simulation, we set $\delta = 8$ dB for shadowing and $n = 4$ for path loss exponent.

We set the threshold power as -56.1236 dBm and plot how the confidence probability varies with distance in Figure 6. It is shown that the confidence probability decreases with distance. To achieve 0.9 confidence level, the CR has to locate closer than 45 m from the PU. Without fading, however, the CR can locate at far as $d = 80$ m to receive -56.1236 dBm power. Therefore, the CR at $d = 45$ m under fading condition should use the same window size as the CR at $d = 80$ without fading. We also try various threshold power and identify the distances for no fading scenario and fading scenario which use the same window size to achieve similar detection performance. The results are shown in Table 2 under 0.9 confidence level. It is shown that the fading has negative impacts for the detection performance.

4.3. Weighted K out of N Fusion Rule. In this section, we compare our proposed weighted K out of N fusion rule with the original K out of N fusion rule. From [23], we know that K is usually chosen as $N/2$ to minimize the total error probability, shown in Figure 7.

We randomly generate $N = 100$ CRs according to a uniform distribution at the distance from the PU ranging

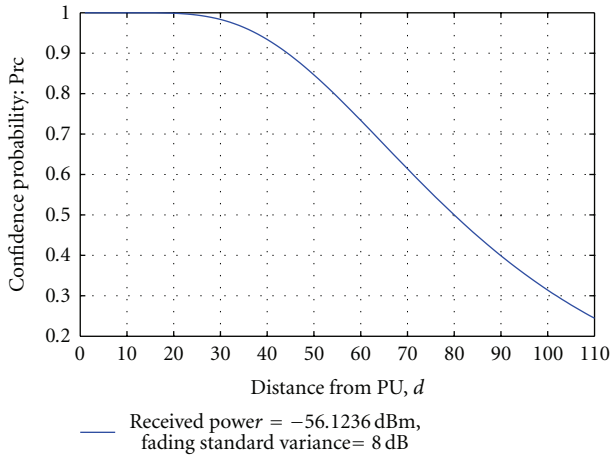


Figure 6: Confidence probability under threshold −56.1236 dBm versus Distance d.

Table 2: Comparing the distance from PU under fading and no fading.

Threshold power (dBm)	No fading	Slow fading
−56.1236	80 (m)	45 (m)
−57.3799	86 (m)	47 (m)
−59.1089	95 (m)	52 (m)
−60.6813	104 (m)	57 (m)
−61.6557	110 (m)	62 (m)

from 80 m to 110 m. We use the linear weight function to assign the weight to each CR. For fair comparison, we set the values A and B such as the expectation of weights is equal to 1 but the variance of the weights can vary. Figures 8 and 9 show the weight assigned for the CRs at different distance with the mean weight as 1 but the weight variance as 0.5 and 0.02, respectively. It is shown that the nearby CR is assigned with higher weight compared to the far-away CR due to the good channel condition. In addition, the weights for CRs at different distances differ more when the weight variance is larger.

We compare the minimum number of individual decisions needed to reach the global decision. We pick $K = N/2 = 50$ so that the total error probability can be minimized in the original K out of N Fusion Rule. For the original fusion rule, each individual decision is treated the same with weight 1, therefore, minimum 50 individual detection decisions with all positive detections are needed. In the weighted fusion rule, the individual decision is treated differently. To minimize the number of individual decisions, we need to include the CRs with the best channel conditions (i.e., the CRs closest to the PU). Example results are shown in Table 3. It is shown that the weighted fusion rule needs less minimum number of individual decisions. In addition, the more the weight variance is, the less number of individual decisions the fusion needs. Since those decisions come from the CRs with best channel conditions and consequently arrive at the central control office quickest, the global decision can be reached quickly without waiting for more decisions.

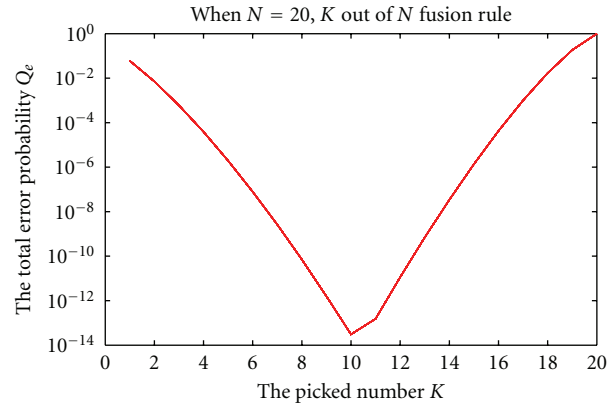


Figure 7: Total error probability versus value K when $N = 20$.

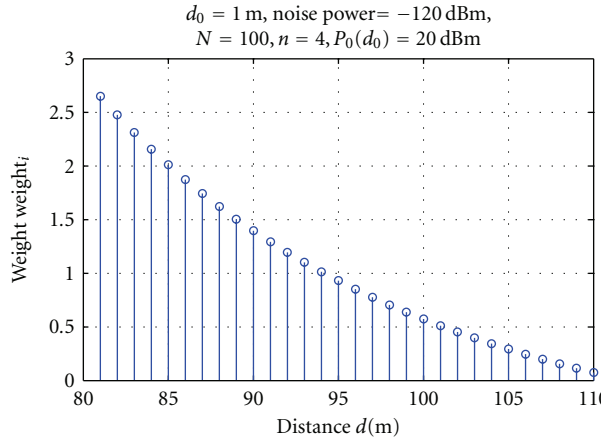


Figure 8: Assigned weight for CRs at different distance when weight variance = 0.5.

Table 3: Comparing the original and the weighted fusions.

$N = 100$	Original fusion	Weighted fusion
variance = 0	$M = 50$	$M = 50$ (same as the original)
variance = 0.02	$M = 50$	$M = 44$
variance = 0.5	$M = 50$	$M = 26$

5. Conclusion

In this paper, we have proposed a fast detection scheme, SPRT-TW for individual detection and weighted K out of N fusion rule for global detection, for cooperative cognitive radio networks. It is shown that the proposed SPRT-TW takes the least detection time compared with traditional NP detection method and the original SPRT and the weighted fusion rule in general takes less numbers of individual decisions (consequently faster) to reach the global decision compared to the original fusion rule. Our scheme takes into consideration the characteristic of wireless channels. For the

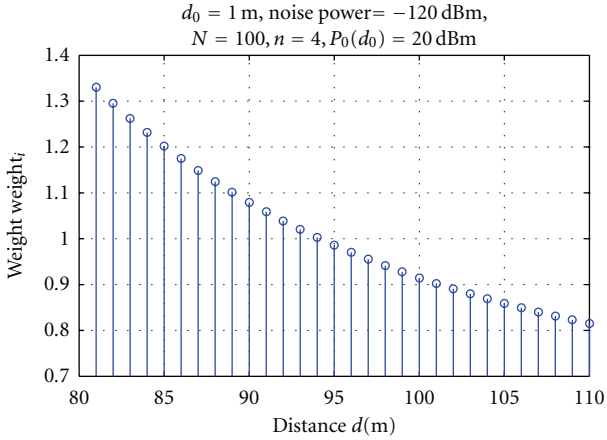


FIGURE 9: Assigned weight for CRs at different distance when weight variance = 0.02.

future work, we will try to derive the optimal design for the weight assignment.

Acknowledgments

This paper was supported by AFOSR under Grant FA9550-09-0630 and by NSF under Grant CNS 0619693.

References

[1] J. Mitola, *Cognitive Radio*, Licentiate thesis, KTH, Stockholm, Sweden, 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] M. A. Sturza and F. Ghazvinian, "White Spaces Engineering Study: can cognitive radio technology operating in the TV white spaces completely protect licensed TV broadcasting?" New America Foundation Wireless Future Program, working paper no.16, January 2007.

[4] K. Watanabe, K. Ishibashi, and R. Kohno, "Performance of cognitive radio technologies in the presence of primary radio systems," in *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, September 2007.

[5] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proceedings of the of the 38th Asilomar Conference on Signals, Systems and Computers*, pp. 772–776, November 2004.

[6] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," in *Proceedings of the International Conference on Communications (ICC '03)*, vol. 5, pp. 3575–3579, May 2003.

[7] E. C. Y. Peh, Y.-C. Liang, and Y. L. Guan, "Optimization of cooperative sensing in cognitive radio networks: a sensing-throughput tradeoff view," in *Proceedings of the IEEE International Conference on Communications (ICC '09)*, pp. 1–5, June 2009.

[8] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 137–143, November 2005.

[9] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[10] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio—part I: two user networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2204–2213, 2007.

[11] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio—part II: multi-user networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2214–2222, 2007.

[12] G. Ganesan and Y. Li, "Agility improvement through cooperative diversity in cognitive radio," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 5, pp. 2505–2509, 2005.

[13] Z. Jiang, X. Zhengguang, W. Furong, H. Benxiong, and Z. Bo, "Double threshold energy detection of cooperative spectrum sensing in cognitive radio," in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '08)*, pp. 1–5, 2008.

[14] J.-W. Ho, M. Wright, and S. K. Das, "Fast detection of replica node attacks in mobile sensor networks using sequential analysis," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 1773–1781, June 2009.

[15] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, 2008.

[16] L.-H. He, X.-Z. Xie, X.-T. Dong, and T. Zhou, "Twice-cooperative spectrum sensing in cognitive radio systems," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, 2008.

[17] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, 2009.

[18] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing in cognitive radio networks," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 131–136, November 2005.

[19] RFC 3561: Ad hoc On-Demand Distance Vector (AODV) Routing, The Internet Society, July 2003.

[20] J. Neyman and E. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London*, pp. 289–337, 1933.

[21] K. P. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, NY, USA, 1997.

[22] W. Wenzhong, Z. Weixia, Z. Zheng, and Y. Yabin, "Detection fusion by hierarchy rule for cognitive radio," in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '08)*, pp. 1–5, May 2008.

[23] W. Zhang, R. K. Mallik, and K. B. Letaief, "Cooperative spectrum sensing optimization in cognitive radio networks," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 3411–3415, 2008.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 823494, 9 pages
doi:10.1155/2010/823494

Research Article

Opportunistic Communication (Cognitive Radio) over Primary Discarded Subchannels by Applying a Double Power Distribution

G. A. Medina-Acosta and José A. Delgado-Penín

Signal Theory and Communications Department, Technical University of Catalonia, Building D4 Campus Nord, Jordi Girona, 1-3, 08034 Barcelona, Spain

Correspondence should be addressed to G. A. Medina-Acosta, agni_medina@tsc.upc.edu

Received 31 October 2009; Revised 2 February 2010; Accepted 12 May 2010

Academic Editor: Fred Daneshgaran

This paper proposes the establishment of a simultaneous cognitive radio communication based on a subdistribution of power made over unselected subchannels which were discarded by the primary user through an initial optimal power allotment. The aim of this work is to show the possibility of introducing an opportunistic communication into a licensed transmission where the total power constraint is shared. The analysis of the proposed transmission scheme was performed by considering 128 and 2048 independent subchannels affected by *Rayleigh* fading, over 10,000 channel realizations, and three different signal-to-noise ratios (8 dB, 16 dB, and 24 dB). From the system evaluation it was possible to find the optimal power allotment for the primary user, the subdistribution of power for the secondary user, as well as the attenuation and the capacity per subchannel for every channel realization. Moreover, the *PDF* and *CDF* of the total obtained capacities, as well as the generation of empirical capacity regions, were estimated as complementary results.

1. Introduction

The radio signals propagating through the environment are associated to a specific operation frequency belonging to one of the many wireless communications systems (i.e., LTE, WiMAX, etc.) existing today, which are strictly allocated by government agencies (i.e., FCC) or international organizations (i.e., ITU) [1, 2]. However, due to the continuous growing and development of the wireless industry, the current static frequency allocation has led to a problem related with spectrum scarcity [3, 4]. Nevertheless, recent worldwide measurement studies have revealed that most of the license spectrum experiences low utilization efficiency [5–7], which means that there exists the possibility of exploiting the underutilized spectrum in an opportunistic manner. According to this, an emerging technology that is able to reliable sense the spectral environment over a wide band, detect the presence/absence of licensed users (primary users), and use the spectrum only if the communication does not interfere with primary users is defined by the term cognitive radio (CR) [8, 9]. So, the spectrum utilization can be improved by making a secondary user access into the spectrum holes or spectrum portions that in a particular location and time are not being used by a primary user. In this regard, according to the current proposals of the CR protocol, the device is constantly aware of its wireless environment in order to determine (at least in space and time) which part of the spectrum is not being occupied by making use of spectrum sensing techniques [10, 11] to later on adapt its signal to fill those spectrum gaps. On the other hand, recent studies have opened the possibility of allowing secondary users to transmit simultaneously with the primary users over the same frequency band [12, 13]. In [14], a proposal that is closely related to the one here described in terms of power allocation involving a primary and a secondary user is presented. Nevertheless, between the main differences it is possible to highlight that the channel characteristics on that research work are considered to be nearly the same for both the primary and the secondary users, while the idea behind the establishment of a simultaneous cognitive radio transmission is based on the assumption that the primary user in any case will not

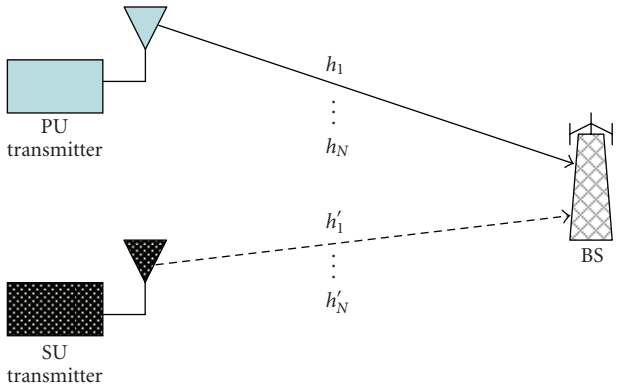


Figure 1: Proposed transmission scheme for establishing a simultaneous low-priority communication with a primary user.

need all its achievable rate; reason why the use of a virtual threshold (target rate) which overestimates the ambient noise is proposed by the authors.

So, by following this last research line, this paper proposes a methodology for establishing a simultaneous low-priority communication (in an opportunistic manner) where the total power constraint is shared. In which, a subdistribution of power is made over small transmission bandwidths (subchannels) that were identified as unselected once an initial optimal power allotment for the primary user (PU) took place. For simulation purposes it was considered that the operation frequency belonging to the primary user was divided in 128 and 2048 independent subchannels of 1 Hz each, affected by *Rayleigh* fading over 10,000 channel realizations and three different signal-to-noise ratios (8 dB, 16 dB, and 24 dB), having destined 20% of the total available power to the secondary user (SU). Among the obtained results, the optimal power allotment for the primary user, the subdistribution of power for the secondary user, as well as the computation corresponding to the attenuation and the capacity per subchannel for one of the channel realizations are shown. Being computed, in addition are the PDF, CDF, and a set of empirically generated capacity regions.

2. Transmission Scheme

The context of this proposal has to do with OFDM systems, being the particular case of a single user OFDM system the candidate scenario for the application of this methodology [15, 16]. So, the PU and SU are assumed to utilize exactly the same harmonic related frequencies (which are carefully chosen to be orthogonal), being the total available transmission bandwidth equally divided amongst N narrowband subcarriers [17–19]. In Figure 1, an uplink considering the presence of a PU and an SU communicating with a base station (BS) over the same frequency band is shown.

Figure 1 is possible to observe a simultaneous communication carried out by two users and a single base station (which is assumed to be equipped with cognitive radio capabilities), where PU is the high-priority user which undergoes several fadings ($h_1 \cdots h_N$) when it is considered

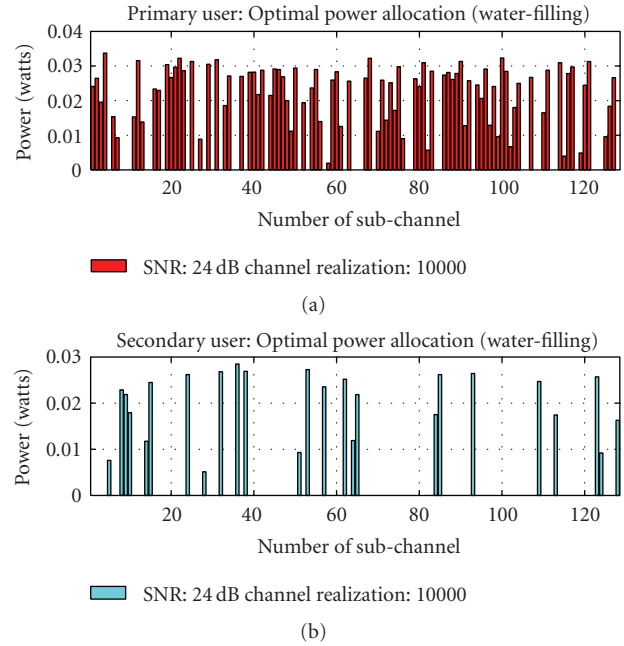


Figure 2: Optimal distribution and subdistribution of power (primary/secondary user), for channel realization number 10,000 with an SNR =24 dB.

that the frequency band over which it operates has been divided in N subbands, same bands that are used for establishing an opportunistic low-priority communication by a secondary user (SU) which undergoes its own fadings ($h'_1 \cdots h'_N$) due to its different spatial location.

So, by assuming that there are N subchannels, the proposal initially focuses on the primary communication over which the mathematical algorithm known as *waterfilling* is applied [20]. The algorithm allows distributing the total available power in an optimal way, which originates that sometimes such as optimization concentrates the power only over certain subchannels (those with the lowest noise levels) leaving the others without using, which corresponds to frequency subbands that can be utilized to transmit low-priority information. Thus, in order to use the spectrum in a more efficient way what is proposed here is to destine a certain percentage (i.e., 20%) of the total available power to a secondary user, so that once the optimal power allotment for the primary user was applied, such information can be used to perform a second power distribution (a new application of the waterfilling algorithm) only over the unused subchannels. Which in agreement with what is presented here turns out to be more efficient in terms of spectrum usage, increasing the total system capacity with respect to the total obtained capacity when the 100% of the total available power is destined to the licensed user.

3. Optimal Power Allotment

The mathematical algorithm known as *waterfilling*, as it was stated before, allows distributing the total available power maximizing the capacity [21, 22]. The algorithm uses the

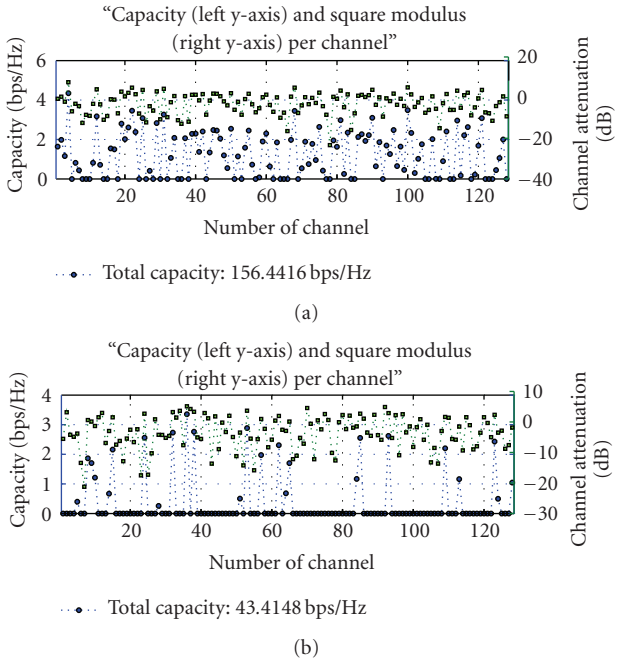


FIGURE 3: Attenuation and capacity per subchannel (primary/secondary user), for the channel realization number 10,000 with SNR = 24 dB.

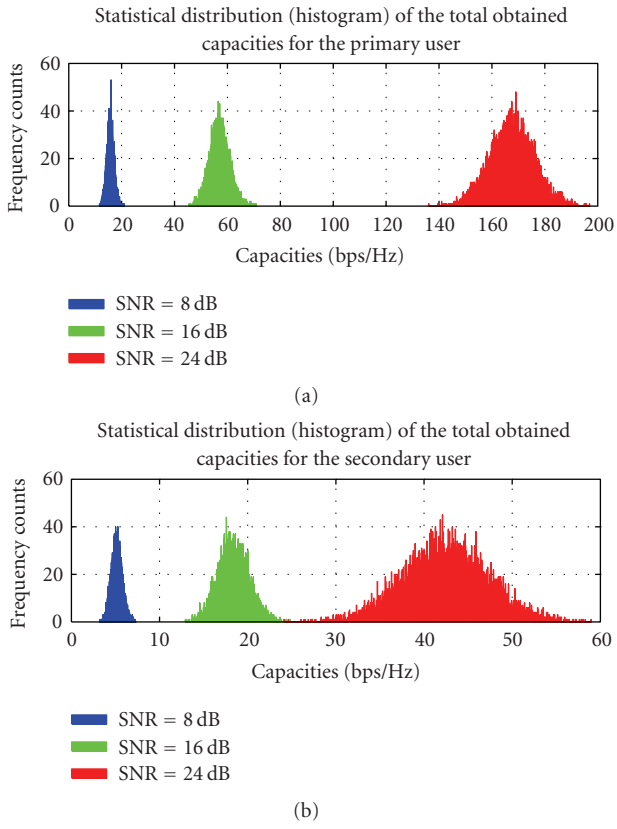


FIGURE 4: Histograms of the total obtained capacities (primary/secondary user), SNR = 8 dB, 16 dB, and 24 dB.

equation corresponding to the total capacity for the primary user (2) subject to a total power constraint (3), leading to an optimization problem that can be solved by using the *Lagrange* multipliers [23].

$$\begin{aligned} \nabla f(x) - \lambda \nabla g(x) &= 0, \\ \nabla F(x) &= 0, \\ \frac{\partial F(x)}{\partial x} &= 0, \end{aligned} \tag{1}$$

where $f(x)$ is the function from which we want to find the extreme values subject to the constraint given by $g(x)$, while λ is known as *Lagrange* multiplier

$$C = \sum_{i=1}^{N} \log\left(1 + \frac{P_i |hi|^2}{\sigma^2}\right) \text{(bps/Hz)}. \tag{2}$$

In this regard the above expression refers to $f(x)$, the total capacity of the primary user which we want to maximize by considering N subchannels, with σ^2 being the noise power, $|hi|^2$ the attenuation, and Pi the power of the i-*esim* subchannel, respectively. While the power constraint represents $g(x)$, which is defined as

$$\sum_{i=1}^{N} P_i = P_{\text{PU}} \text{ (watts)}. \tag{3}$$

Which establishes that the sum of each of the powers assigned to the subchannels must be equal to a total given available power (in this case the one corresponding to the primary user P_{PU}). So, based on the two above equations and by using the *Lagrange* multipliers it is possible to write,

$$F(P_i) = \sum_{i=1}^{N} \log\left(1 + \frac{P_i |hi|^2}{\sigma^2}\right) - \lambda\left(\sum_{i=1}^{N} P_i - P_{\text{PU}}\right). \tag{4}$$

The above expression corresponds to a function which, when is derived with respect to P_i, allows to find the optimal power distribution

$$\frac{1}{\lambda} - \frac{\sigma^2}{|hi|^2} = P_i \text{ (watts)}. \tag{5}$$

So, by following what is established in the above equation, in order to optimally allocate the power for every subchannel it is necessary to previously determine the value for the constant $1/\lambda$, which corresponds to a level of power that acts as a threshold and that is defined by,

$$\frac{1}{\lambda} = \frac{P_{\text{PU}} + \sum_{i=1}^{N} \sigma^2/|hi|^2}{N}. \tag{6}$$

Which can be found by substituting the result obtained in (5) directly in (3). On the other hand, because the algorithm could assign negative powers, it is necessary to

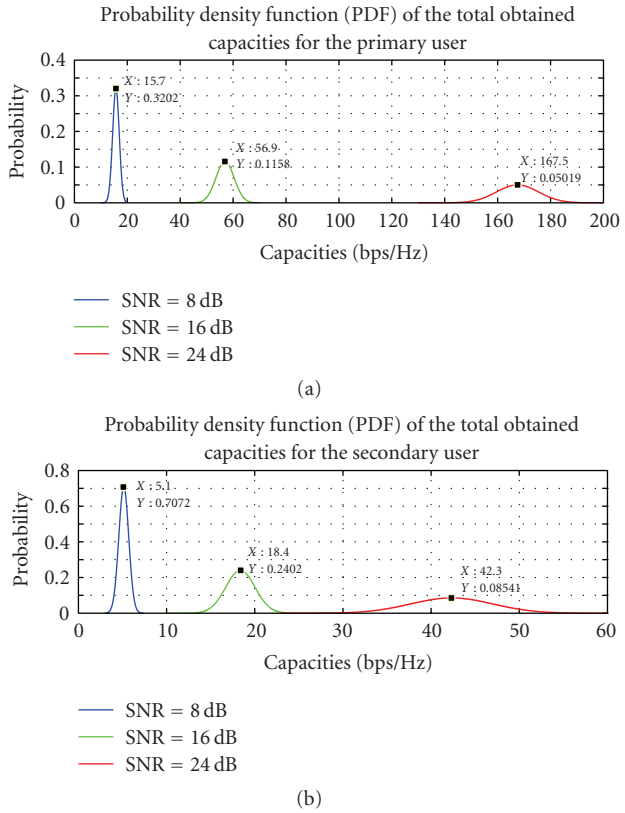


FIGURE 5: PDFs of the total obtained capacities (primary/secondary user), SNR = 8 dB, 16 dB, and 24 dB.

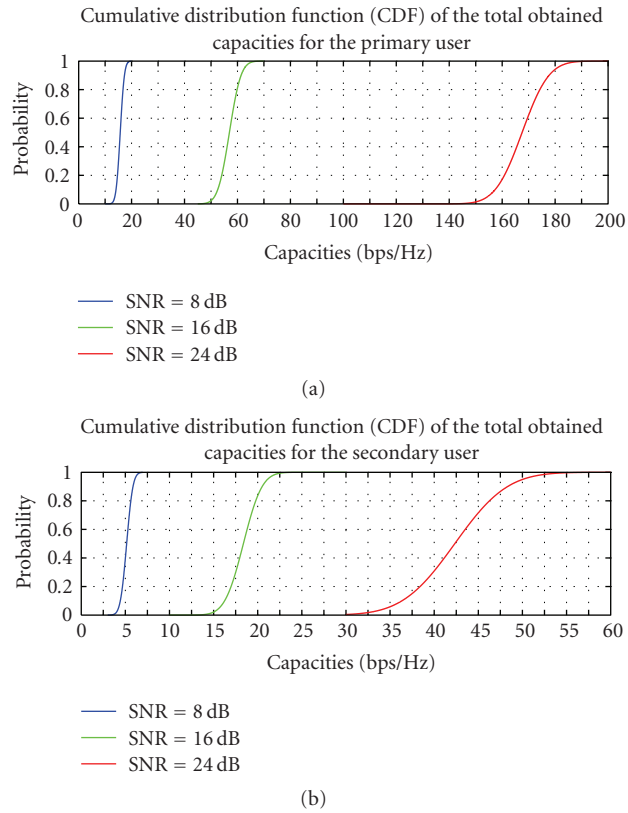


FIGURE 6: CDFs of the total obtained capacities (primary/secondary user), SNR = 8 dB, 16 dB, and 24 dB.

apply the conditions known as *Kuhn-Tucker* [24]

$$[x]^{+} = \begin{cases} x & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{7}$$

In this context, after considering that P_i takes the value of x, the imposed conditions are established to assign a zero whenever a negative power is obtained. Consequently it is necessary to find a new threshold and to recalculate the power allotment by only taking into account those channels for which we previously had obtained positive powers, discarding therefore the rest. This concludes the description of the methodology that allows optimally allocating the available power for the primary user. Hence, the following step consists in describing the insertion of a secondary user into the system by considering the implementation of a strategy related with a subdistribution of power.

4. Subdistribution of Power

According to what is proposed here, a certain percentage of the total system's available power (P) would be destined to a secondary user, therefore $P = P_{\text{PU}} + P_{\text{SU}}$. Being necessary to take into account that the primary user must not be affected by the presence of the low-priority user is the reason why such communication would not have to notice the establishment of a secondary communication in

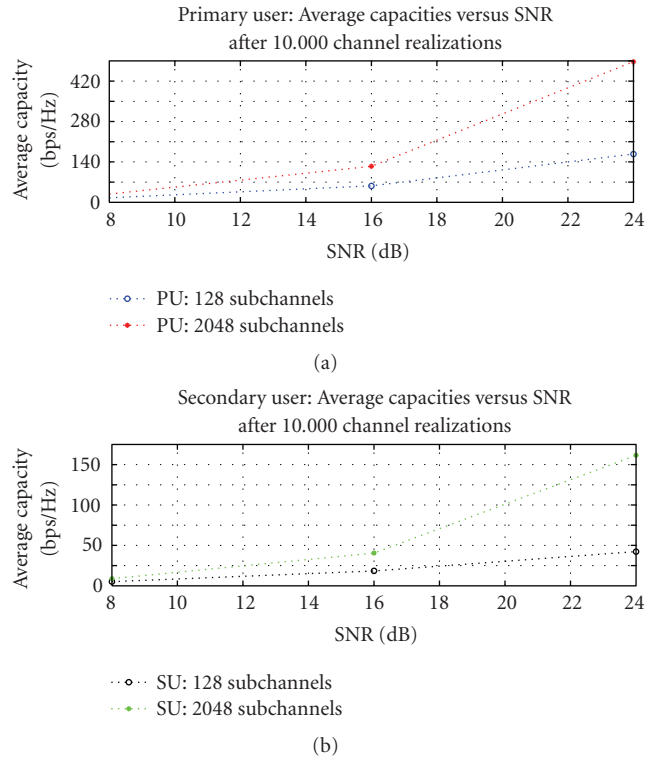


FIGURE 7: Average Capacities comparison for 128 and 2048 subchannels (primary/secondary user), SNR = 8 dB, 16 dB, and 24 dB.

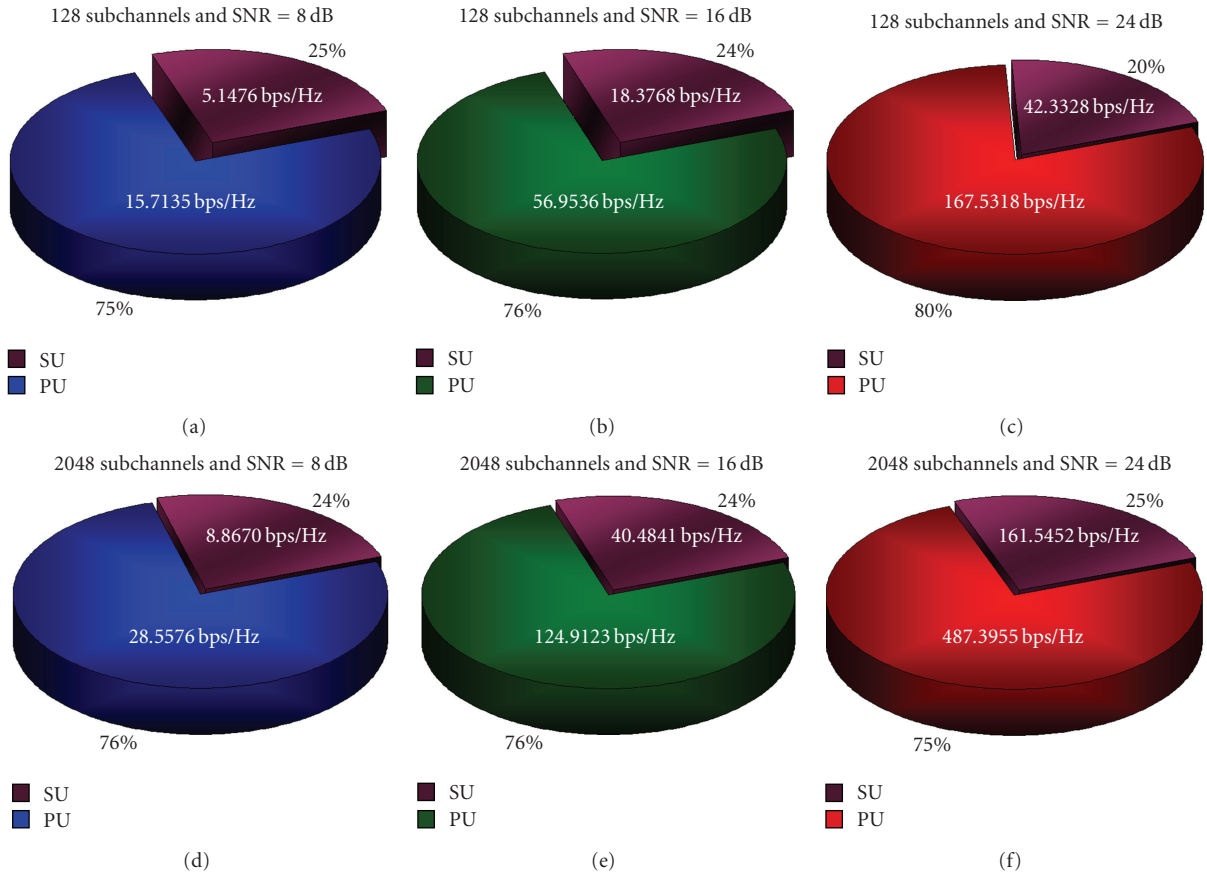


Figure 8: Percentage contribution (PU and SU) in terms of the average capacities (128/2048 subchannels), SNR = 8 dB, 16 dB, and 24 dB.

parallel, while this last one would perceive the high-priority communication as noise. This leads to the following equation corresponding to the total capacity for the secondary user:

$$C' = \sum_{i=1}^{N} \log\left(1 + \frac{P'_i |h'_i|^2}{P_i |h_i|^2 + \sigma^2}\right) \text{(bps/Hz)}, \tag{8}$$

where the terms $|h'_i|^2$ and P'_i correspond to the attenuation and the power of the *i-esim* subchannel for the secondary user, respectively. For its part, the power constraint in this case is defined as

$$\sum_{i=1}^{N} P'_i = P_{\text{SU}} \text{ (watts)}. \tag{9}$$

Which infers that the sum of the powers allocated to the subchannels must be equal to the percentage of power destined to the secondary user. So, based on the above expressions it is possible to apply once again the algorithm known as *waterfilling*, which together with the *Kuhn-Tucker* conditions produces as optimal solution the following:

$$\left[\frac{1}{\lambda} - \left(\frac{P_i |h_i|^2 + \sigma^2}{|h_i|'^2}\right)\right]^+ = P'_i \text{ (watts)}. \tag{10}$$

It is once again necessary to determine the value corresponding to the constant $1/\lambda$, which can be found by substituting (10) in (9)

$$\frac{1}{\lambda} = \frac{P_{\text{SU}} + \sum_{i=1}^{N} \left(P_i |h_i|^2 + \sigma^2\right) / \left(|h_i|'^2\right)}{N}. \tag{11}$$

So, under these considerations it is possible to establish a simultaneous low-priority communication over the same operation band that originally is utilized only by the licensed user.

Summarizing the proposal, in general once the optimal power allotment for the primary user applied it was possible to know which subchannels were discarded by the optimization algorithm, or in other words to which of them was power not given. Thus, by making use of this information it is possible to apply a second optimization algorithm only over those subchannels which were previously indentified as available, leading to a subdistribution of power that concentrates P_{SU} exclusively on those subbands. Reason why the low-priority communication is established in an opportunistic manner over the unused subchannels is avoiding, this way, to interfere with the primary user, whereas at the same time the spectrum is utilized in a more efficient way.

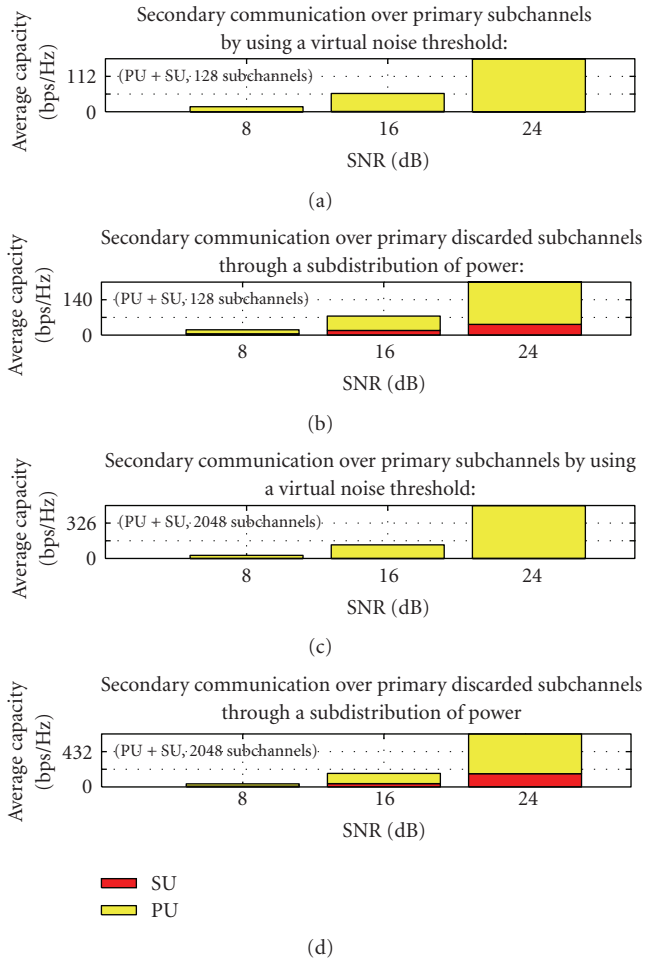


Figure 9: System's performance comparison (128/2048 subchannels, $P_{PU} = 80\%$ & $P_{SU} = 20\%$), SNR = 8 dB, 16 dB, and 24 dB.

5. Results

In this section, the obtained results after applying the methodology described before are discussed. In Figure 2, the optimal power allotment for the primary user (PU), as well as the subdistribution of power by the secondary user (SU), for one of the channel realizations (number 10,000/10,000) considering a signal-to-noise ratio equal to 24 dB and when the total transmission bandwidth is equally divided in 128 subchannels is shown.

In Figure 2(a) the result corresponding to the application of the first optimization algorithm (*waterfilling*) is shown, where it is possible to observe that some of the subchannels were discarded or not taken into account for establishing the primary communication (there is no power allocated to these subchannels). Which, as can be observed in the low part of the figure were utilized for establishing a low-priority communication through an optimal subdistribution of power made exclusively on those subbands.

On the other hand, in Figure 3, for the same channel realization and the same signal-to-noise ratio, the attenuation and the obtained capacity for each of the subchannels are shown.

Figure 3 is possible to note that for the first case (primary user), the subchannels having the best channel conditions reached higher capacities while those that presented a more severe attenuation were discarded, this is the reason why its capacity is equal to zero (as it happens with the subchannels 5, 62, 105, and 122 to mention some). On the other hand, for the second case (secondary user) it is possible to verify that the unused frequency subbands by the primary user were seen as available by the secondary user, who undergoes different channel conditions and that after applying an optimal subdistribution of power was able to obtain (for most of the cases and depending on the attenuation) capacities different from zero on these subchannels, which allows establishing a low-priority communication in parallel. In Figure 4, the statistical distribution of the total obtained capacities for each of the signal-to-noise ratios considered in this analysis is shown.

By taking into account all the channel realizations and once the last one took place, it was possible to determine the statistical distribution (*Histogram*) of the total obtained capacities. Where it can be verified that the total obtained average capacities PU: 15.7135 bps/Hz, 56.9536 bps/Hz, and 167.5318 bps/Hz / SU: 5.1476 bps/Hz, 18.3768 bps/Hz, and 42.3328 bps/Hz are located (resp.) on the intervals corresponding to the highest point in each of the histograms. Additionally, by making use of maximum likelihood estimators [25–27], it was found that the data corresponding to the total obtained capacities fit better to normal distributions with parameters shown in Table 1.

In Figures 5 and 6, respectively, approximations for the probability density function (PDF) and cumulative distribution function (CDF) are shown.

This way, from the approximations made for the PDF and CDF it is possible to extract valuable information related with the probability of observing a certain capacity. For example when as SNR = 24 dB, for the PUs the probability of observing a capacity less than or equal to 165.6 bps/Hz is 0.404 while for the SU there is a probability equal to 0.4042 of observing a capacity less or equal to 41.2 bps/Hz.

On the other hand, aiming at extending the results an analogue procedure was followed in order to obtain the average capacities by considering now 2048 subchannels, which, in Figure 7, were compared to the case of 128 subchannels.

Figure 7 shows a comparison between the previously discussed average capacities for the case of 128 subchannels and the obtained ones for the case of considering 2048 subchannels (PU: 28.5576 bps/Hz, 124.9123 bps/Hz, and 487.3955 bps/Hz / SU: 8.8670 bps/Hz, 40.4841 bps/Hz, and 161.5452 bps/Hz). Where it is possible to observe that for most of the cases (considering both 128 & 2048 subchannels) the SU reaches average capacities about one third with respect to those ones obtained by the PU.

Moreover, when the primary user (PU) and the secondary user (SU) are seen as unique whole system (which can be assumed since they share the total available power), their percentage contributions in terms of average capacities for each of the SNRs considered before are shown through a pie chart in Figure 8.

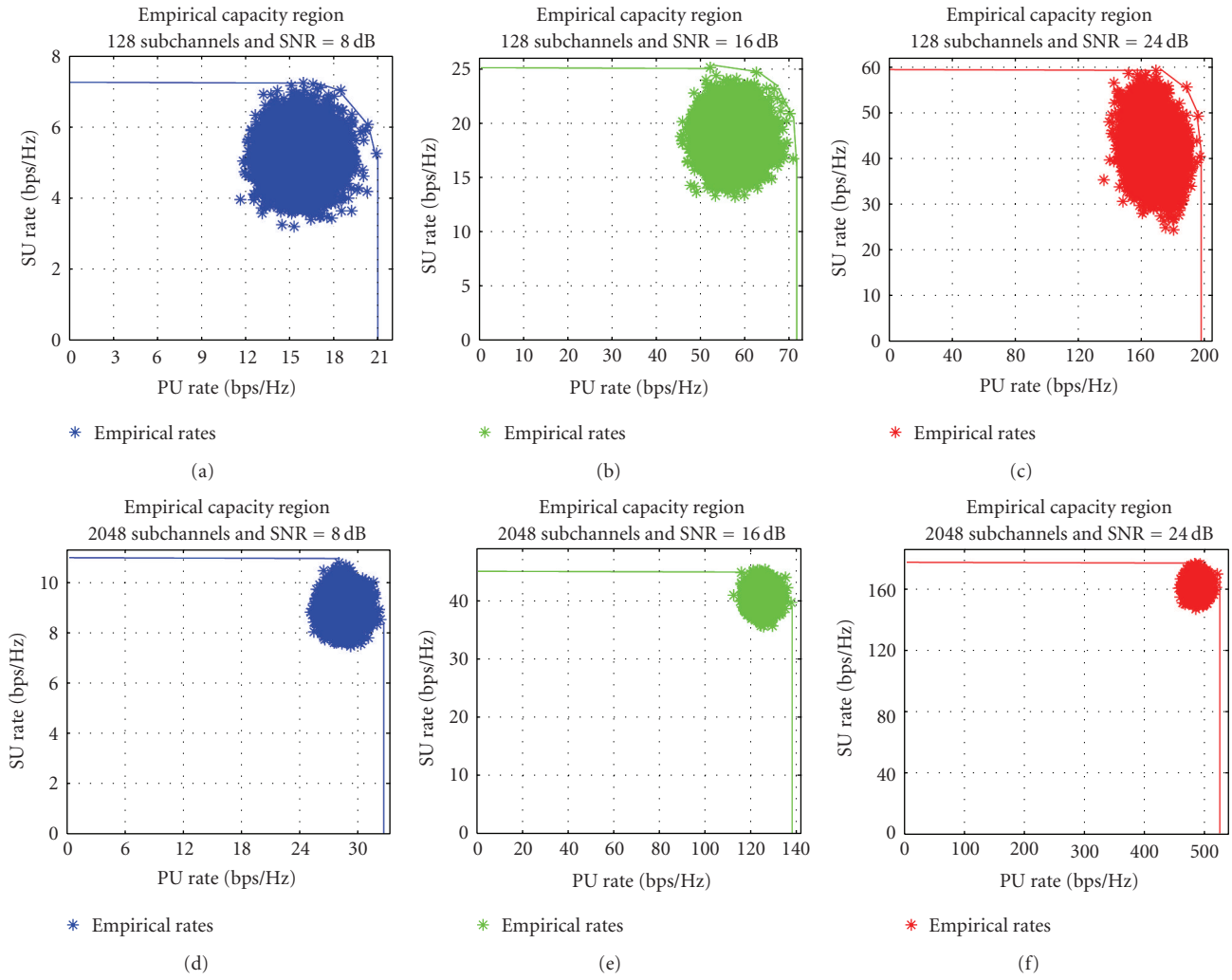


Figure 10: Empirical capacity regions (128/2048 subchannels, $P_{PU} = 80\%$ & $P_{PU} = 20\%$), SNR= 8 dB, 16 dB, and 24 dB.

Table 1: Parameters of the fitted curves.

SNR	PU	SU
8 dB	$\mu = 15.7135$	$\mu = 5.14764$
	$\sigma^2 = 1.55186$	$\sigma^2 = 0.315983$
16 dB	$\mu = 56.9537$	$\mu = 18.3768$
	$\sigma^2 = 11.8711$	$\sigma^2 = 2.75803$
24 dB	$\mu = 167.532$	$\mu = 42.3328$
	$\sigma^2 = 63.1848$	$\sigma^2 = 21.8141$

Figure 8 allows visualizing the percentage contributions that the PU and SU make when they are considered as a single system. Under this assumption (PU+SU), the total average capacities supplied by the system turned out to be 20.8611 bps/Hz, 75.3304 bps/Hz, and 209.8646 bps/Hz & 37.4246 bps/Hz, 165.3964 bps/Hz, and 648.9407 bps/Hz for 128 and 2048 subchannels, respectively.

In this context, and in order to compare the obtained results, the principles stated in the proposal cited in [14], about considering the same channel impairments for PU & SU as well as the use of a given virtual noise threshold (chosen in this case arbitrarily to overestimate the ambient noise by 35%), were simulated under similar conditions (number of channel realizations, signal-to-noise ratios, and power sharing scheme) to those used in our analysis, whose results are shown in Figure 9.

After comparing the results, (for most of the signal-to-noise ratio levels considered) a low (even null) performance by part of the SU in the case of the proposal based on the virtual noise threshold can be noted, this due to

the inherent differences between the foundations of the proposals, because while the first one originally treats the PU & SU as independent entities in terms of technical resources (i.e., no power sharing is contemplated, and therefore more power is required for its proper operation), the second one (our proposal) aims to maximize the use of the already available resources, which is fulfilled after observing the significant contribution made by the SU to the system's throughput.

On the other hand, when the same number of channel realizations is considered by only taking into account the primary user (that means by assigning all the available power to PU) the average capacities turned out to be 18.5231 bps/Hz, 65.7161 bps/Hz, and 187.6908 bps/Hz for the case of 128 subchannels & 34.3087 bps/Hz, 148.3475 bps/Hz, and 568.8980 bps/Hz for the case of 2048 subchannels, which do not differ in a significant way from the obtained capacities when the power is shared with a secondary user, and nevertheless if these last ones (secondary) are grouped with the primary ones (as we verified previously in the case of our proposal), they exceed all the total obtained capacities on average. This is reason why we use an opportunistic subdistribution of power over primary discarded subchannels for establishing a low-priority communication, which results in a more efficient usage of the available (natural & technical) resources.

Finally, bidimensional surfaces were generated in order to provide empirical approaches of the system's capacity region for each of the signal-to-noise ratios considered above, which are shown in Figure 10.

Figure 10 highlights the empirically obtained rate combinations (i.e., 128 subchannels & 8 dB: 20.89 bps/Hz, 5.255 bps/Hz / 2048 subchannels & 24 dB: 521.2 bps/Hz, 169.4 bps/Hz) for each of the 10,000 channel realizations, which correspond to points located inside a region representing the system's achievable rates (error-free) when it is considered that the PU & SU transmit simultaneously.

The capacity regions for Cognitive Radios have been theoretically studied and developed in [28–30], the results here being obtained nearer to be classified as an hybrid scheme that combines the overlay-interweave paradigm for Cognitive Radios since a power split is utilized, because the SU utilizes knowledge about the PU channel conditions (discarded subchannels) and because of the interference mitigation given by the orthogonal nature of this proposal. In concrete from the empirically obtained results, it can be verified that in spite of the rate reduction produced sometimes by having only a few available (primary discarded) subchannels, the SU was able to achieve (for each of the 10,000 channel realizations) opportunistic rates, in parallel, different from zero, optimizing, this way, the spectrum usage.

6. Conclusions

In this paper, a strategy for establishing a secondary communication where the total system's available power is shared involving a double optimization procedure is proposed. On the basis of a scenario where the operation band belonging to the primary user is equally divided in several (128 & 2048) subbands, and once an optimal power distribution is applied by using the algorithm known as *waterfilling*, the proposed methodology (taking advantage of the system's orthogonality) consists in making use of those subchannels that were not used or discarded by the PU for applying a subdistribution of power leading to the establishment of a low-priority communication in parallel. From the analysis of the obtained results it was possible to find approximations for the PDF and CDF in both cases PU & SU and to later on verify that despite of having presence in an opportunistic way the secondary user could reach average capacities up to one third of the obtained ones by the primary user, which turns out to be useful if we consider that initially those resources are not used. Moreover, if the primary and the secondary users are seen as a unique system, they exceed the total average capacities with respect to those obtained when a conventional single primary transmission (using exactly the same resources) is considered. In addition, the rate combinations obtained for each of the 10,000 channel realizations were utilized in order to create an approach about the system's achievable rates through the generation of empirical capacity regions. On the other hand, in terms of scalability it can be inferred that the fact of using a methodology like this could allow incorporating a low-priority communication per every licensed user located in an existing system (i.e., single user OFDM system). So, in order to finalize it is possible to conclude that by destining a certain percentage of the total available power to a secondary user and by applying a strategy related with a double distribution of power as suggested here, it is possible to use the spectrum in more efficient way without interfering and modifying drastically the obtained capacities when the total available power is destined only to a primary user.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments.

References

[1] Federal Communications Commission (FCC), http://www.fcc.gov.

[2] The ITU Radio Communications Sector (ITU-R), http://www.itu.int/ITU-R.

[3] R. W. Brodersen, A. Wolisz, D. Cabric, S. M. Mishra, and D. Willkomm, "A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum," UC Berkeley White Paper, July 2004.

[4] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 151–159, November 2005.

[5] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, pp. 772–776, November 2004.

[6] Shared Spectrum Company, "Recent Measurements," http://www.sharedspectrum.com/measurements/recent.html.

[7] M. H. Islam, L. K. Choo, W. O. Ser et al., "Spectrum survey in Singapore: occupancy measurements and analyses," in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '08)*, May 2008.

[8] D. Čabrić and R. W. Brodersen, "Physical layer design issues unique to cognitive radio systems," in *Proceedings of the 16th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '05)*, vol. 2, pp. 759–763, September 2005.

[9] FCC, "Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies," ET Docket 03-108, December 2003.

[10] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[11] H. Arslan, *Cognitive Radio Software Defined Radio and Adaptive Wireless Systems*, Springer, Berlin, Germany, 2007.

[12] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, 2006.

[13] A. Jovičić and P. Viswanath, "Cognitive radio: an information-theoretic perspective," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '06)*, pp. 2413–2417, Seattle, Wash, USA, July 2006.

[14] M. Haddad, A. M. Hayar, and M. Debbah, "Optimal power allocation for cognitive radio based on a virtual noise threshold," in *Proceedings of the 1st International Workshop on Cognitive Wireless Networks (CWNETS '07)*, August 2007.

[15] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171–178, 2003.

[16] M. Saleh, "Adaptive resource allocation in multiuser OFDM systems," Literature Survey on Multidimensional Digital Signal Processing, Texas University, 2005.

[17] T. Hwang, C. Yang, G. Wu, S. Li, and G. Y. Li, "OFDM and its wireless applications: a survey," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1673–1694, 2009.

[18] V. Chakravarthy, A. S. Nunez, J. P. Stephens, A. K. Shaw, and M. A. Temple, "TDCS, OFDM, and MC-CDMA: a brief tutorial," *IEEE Communications Magazine*, vol. 43, no. 9, pp. S11–S16, 2005.

[19] X. Huang and H.-C. Wu, "Intercarrier interference analysis for wireless OFDM in mobile channels," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '06)*, vol. 4, pp. 1848–1853, April 2006.

[20] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[21] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, NY, USA, 1968.

[22] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.

[23] J. Stewart, *Multivariable Calculus: Concepts and Contexts*, Thompson, Boston, Mass, USA, 2004.

[24] S. Haykin, *Communications Systems*, John Wiley & Sons, New York, NY, USA, 1991.

[25] "Likelihood and method of Maximum Likelihood (AI-ACC)," http://www.aiaccess.net/English/Glossaries/GlosMod/e_gm_.. likelihood.htm.

[26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Upper Saddle River, NJ, USA, 1993.

[27] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Texts in Statistics, Springer, New York, NY, USA, 2nd edition, 1998.

[28] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *IEEE Communications Magazine*, vol. 44, no. 6, pp. 44–49, 2006.

[29] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, 2006.

[30] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: an information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 458636, 11 pages
doi:10.1155/2010/458636

Research Article

Resource Allocation with MAC Layer Node Cooperation in Cognitive Radio Networks

Andreas Merentitis and Dionysia Triantafyllopoulou

Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis, Ilissia, 15784 Athens, Greece

Correspondence should be addressed to Dionysia Triantafyllopoulou, siarina@di.uoa.gr

Received 1 November 2009; Revised 29 January 2010; Accepted 21 February 2010

Academic Editor: Hamid Sadjadpour

An algorithm for cooperative Dynamic Spectrum Access in Cognitive Radio networks is presented. The proposed algorithm utilizes Medium Access Control layer mechanisms for message exchange between secondary nodes that operate in license exempt spectrum bands, in order to achieve interference mitigation. A fuzzy logic reasoner is utilized in order to take into account the effect of the coexistence of a large number of users in the interference as well as to cope for uncertainties in the message exchange, caused by the nodes' mobility and the large delays in the updating of the necessary information. The proposed algorithm is applied in Filter Bank Multicarrier, as well as Orthogonal Frequency Division Multiplexing systems, and its performance is evaluated through extensive simulations that cover a wide range of typical scenarios. Experimental results indicate improved behaviour compared to previous schemes, especially in the case of uncertainties that cause underestimation of the interference levels.

1. Introduction

The proliferation of mobile devices, coupled with the ever-increasing demand for higher data rates' support, constitutes static frequency allocation schemes suboptimal, as they frequently result in spectrum underutilization. Cognitive Radios (CRs) supporting Opportunistic Spectrum Access (OSA) [1] emerged as a new paradigm that offers an effective solution to the problem of spectrum scarcity. However, the increased variance in spectrum availability combined with the end users' diverse characteristics and Quality of Service (QoS) requirements poses a number of challenges that need to be addressed.

More specifically, for Cognitive Radio systems operating in licensed spectrum bands with coexistence of both primary and secondary users, the operations of spectrum sensing, defined as the identification of the spectrum bands that are available for transmission, and spectrum mobility, that is, the vacation of the wireless channel when a primary user is detected, are of key importance. On the other hand, Cognitive Radio systems operating in license exempt spectrum bands, where different operators coexist, require efficient spectrum decision and spectrum sharing algorithms as well as power control mechanisms for interference mitigation. For example, if all users transmit at the maximum valid power level, then every user is causing significant interference to all the others, a fact that can result in reduced total utility from the network perspective and, finally, poor QoS for the end users.

In this scope, algorithms that employ cooperative spectrum sharing in order to maximize the overall system performance are required. These algorithms need to be *distributed*, in order to be applied efficiently in ad hoc networks operating in unlicensed spectrum bands where synchronization is necessary only for users of the same operator. Such algorithms should also be able to employ *efficient message exchange* schemes in order to maximize the overall system utility (therefore, the related systems are classified as cooperative CR systems); however, uncertainties in message exchange should also be considered. Furthermore, they should be able to *converge* to an optimal solution within a finite number of iterations to be applicable to real systems.

In order to address some of the previous challenges, the authors in [2] propose a price-based iterative water-filling algorithm which allows users to converge to the Nash Equilibrium. This algorithm can be implemented in a distributed manner with CRs negotiating their best transmission powers and spectrum. In [3], a Dynamic Open Spectrum Sharing Medium Access Control (MAC) protocol for wireless ad hoc networks is proposed. This protocol performs real-time dynamic spectrum allocation by allowing nodes to adaptively select an arbitrary spectrum for the incipient communication subject to spectrum availability. In [4], a distributed approach to spectrum allocation that starts from the previous spectrum assignment and performs a limited number of computations to adapt to recent topology changes is considered. According to the proposed local bargaining approach, the users affected by a mobility event self-organize into bargaining groups and adapt their spectrum assignment to approximate a new optimal conflict-free assignment. The authors in [5] propose a graph-theoretic model for efficient and fair access in open spectrum systems. Three policy-driven utility functions that combine efficient spectrum utilization and fairness are described, and a vertex labeling mechanism is used to build both centralized and distributed approximation algorithms. In [6], a group-based coordination scheme, and distributed group setup and maintenance algorithms where users select coordination channels adaptively are proposed. In [7], an algorithm that allows for transmission power and transmission frequencies to be chosen simultaneously by Cognitive Radios competing to communicate over a frequency spectrum is proposed. Finally, in [8], an algorithm in which each user selects a single channel along with its transmission power by taking into account information concerning the interference caused to other users in the network is introduced.

In this paper, an algorithm based on the spectrum sharing scheme of [8] for distributed interference compensation in Cognitive Radios that operate in license exempt spectrum bands is proposed. The proposed algorithm refines the utility function used in [8] to improve the system scalability in the case of a large number of user pairs and to take into account uncertainties that may be the result of user mobility and large delays in the update of the interference prices. More specifically, a fuzzy logic reasoner is utilized in order to take into account the effect of a large number of users and the related interference as well as to cope for uncertainties in the message exchange process. The performance of the proposed algorithm is evaluated through simulations. In this direction, the overall utility value of the algorithm is compared to the utility of a simple "always select the maximum valid power" policy. The proposed algorithm is also applied in both Filter Bank Multicarrier (FBMC) and Orthogonal Frequency Division Multiplexing (OFDM) systems in order to show its flexibility and capability of transparently exploiting an improved Physical layer, without any further modifications. Moreover, comparison with the distributed algorithm of [8] is used to validate the improvement in terms of the overall utility level under uncertainties that cause 25% underestimation of the interference. Finally, in order to quantify the improvement using conventional network metrics and to show the relation between a higher overall utility value and parameters that directly affect the user experience, comparison with the algorithm of [8] in terms of Signal-to-Interference-plus-Noise Ratio (SINR) is also performed. Experimental results indicate that SINR is consistently improved with the use of the proposed algorithm.

The rest of the paper is organized as follows. Section 2 describes in detail the proposed algorithm. Fuzzy Inference for defining the algorithm parameters is outlined in Section 3. In Section 4, the performance of the proposed algorithm is evaluated through simulations. Finally, Section 5 contains conclusions and plans for future work.

2. Algorithm Outline

The main idea of the proposed algorithm is that users exchange information concerning their interference levels, using explicit MAC layer message exchange mechanisms. A transmitter sets its power level by considering not only its own SINR information, but also the negative impact in utility for other users caused from the greater interference that will come as a side effect of the increase in transmission power of that particular user. This functions as a counter-motive that prevents users from always increasing their transmission power towards the maximum valid level.

Assuming a system with a total of L user pairs in a spectrum band with K available channels, the SINR of the ith user pair, $i \in \{1, 2, \ldots, L\}$ in the kth channel, $k \in \{1, 2, \ldots, K\}$, is given by the following equation [8]:

$$\gamma_i\left(p_i^k\right) = \frac{p_i^k \cdot h_{ii}}{n_0 + \sum_{j \neq i} p_j^k \cdot h_{ji}}, \tag{1}$$

where p_i^k is the transmission power for user i on channel k, h_{ii} is the link gain between the ith receiver and the ith transmitter, $n_0 = 10^{-2}$ is the noise level, p_j^k, $j \in \{1, 2, \ldots, L\}$, $j \neq i$, is the transmission power for all other users on channel k, and h_{ji} is the link gain between the ith receiver and the jth transmitter. It should be noted that $h_{ij} \neq h_{ji}$, since the first expresses the gain between the ith transmitter and the jth receiver and the latter expresses the gain between the jth transmitter and the ith receiver.

In the general case, the carrier frequency of a signal is varied; therefore the magnitude of the change in amplitude will also vary. The coherence bandwidth measures the separation in frequency after which two signals will experience uncorrelated fading. More specifically, in the case of frequency-selective fading, the coherence bandwidth of the channel is smaller than the bandwidth of the signal. Thus, different frequency components of the signal experience decorrelated fading. On the other hand, in the case of flat fading, the coherence bandwidth of the channel is larger than the bandwidth of the signal. Therefore, all frequency components of the signal will experience the same magnitude of fading. In the following analysis, a flat-faded channel without shadowing effects is assumed. For a flat-faded channel, there are no delay spread and no frequency selectivity, as mentioned previously. Thus, a single coefficient

is used for channel attenuation. Since the described channel is static, that is, the coefficient is fixed, the only attenuation present is the path loss. Therefore, in this particular case, h is strictly the channel attenuation or channel gain. In this paper, the environment is assumed to cause average-to-high loss (path loss exponent equals three, a value typical for indoor urban environments), thus the channel gain is $h_{ji} = d_{ji}^{-3}$, where d is the distance between the jth transmitter and the ith receiver.

In order to model the impact on utility for user i caused by the transmission of all other users, the notion of interference price is adopted from [8]. Interference price is defined as

$$\pi_i^k = \frac{\partial u_i\left(\gamma_i\left(p_i^k\right)\right)}{\partial\left(\sum_{j\neq i} p_j^k \cdot h_{ji}\right)}, \tag{2}$$

where $u_i(\gamma_i(p_i^k)) = \theta_i \log(\gamma_i(p_i^k))$ is the logarithmic utility function and θ_i is a user-dependent parameter. As shown, the interference price expresses the marginal utility degradation due to a marginal increase in sustained interference. Interference prices are exchanged between the users in a completely asynchronous fashion, while every user is able to update its own price and power level at different times. Each user selects an appropriate transmission power level in order to maximize the difference between the increase in its own utility minus the utility degradation for others, caused by the increased interference as expressed by the interference price. Specifically, the mathematical formula that [8] attempts to maximize is

$$u_i\left(\gamma_i\left(p_i^k\right)\right) - p_i^k \sum_{j\neq i} \pi_j^k \cdot h_{ji}. \tag{3}$$

The first part of this equation is closely related to the Shannon capacity for user i (the constant term is excluded in order to have a form that can be proved to converge in all cases [8]). Increasing that part is directly related to an increase in the maximum bit rate. However, since the transmission of every user is considered as noise by the other users, the second term expresses the utility loss of the other users if user i increases its transmission power level.

The algorithm is comprised by the following steps.

(1) *Initialization*: For every user $i \in \{1, 2, \ldots, L\}$ transmitting in channel k, select a valid transmission power level p_i^k and a positive value for the interference price π_i^k.

(2) *Power Update*: For every user i at a time interval $t_{ai} \in T_i$, where T_i is a set of positive time instances in which the user i will update its transmission power level and $t_{a1} \neq t_{a2} \neq \cdots \neq t_{ai}$, set p_i^k to maximize (3).

(3) *Interference Price Update*: For every user i at a time interval $t_{bi} \in T_i'$, where T_i' is a set of positive time instances in which the user i will update its interference price and $t_{b1} \neq t_{b2} \neq \cdots \neq t_{bi}$, calculate and announce the updated interference price π_i^k and notify the other users for the updated value.

Steps (2) and (3) are repeated asynchronously for all users until the algorithm reaches its final steady state. In order to perform the power update of step (2), users select p_i^k from the set TP of the allowable transmission power levels, so that the surplus of (3) is maximized. Provided that the allowable power levels are equidistant values with each one being derived from its previous by adding a constant increment, then it can be proved that the algorithm converges, as long as the increment is sufficiently small. Moreover, if the problem is partitioned so that there is a single available spectrum area, or the algorithm is executed only for subgroups selecting the same spectrum area M, then it converges to a global maximum under arbitrary asynchronous updates [8].

In order to execute the algorithm, every user in the network needs to know its own SINR value and channel gain as well as the channel gains and the interference prices announced by other users. The SINR and the channel gain between a user pair can be calculated at the receiver and forwarded back to the transmitter. The channel gains between users can be calculated if receivers periodically broadcast a beacon [9] (h_{ii} message between Receiver i and Transmitter i in Figure 1). This information can also be provided on demand through a specially defined message sent from the receiver. Thus, in case the transmitter requires channel gain information before the reception of the next scheduled beacon, it can request this information from the receiver who will respond with the relative measurements. Finally, interference price values can be also conveyed in the same manner (message π_{ij} from Receiver i to Transmitter j in Figure 1). Every user announces a single interference price, therefore the delay that is introduced by the algorithm scales linearly with the number of users. This also implies that, given the fact that the updates are distributed in an asynchronous manner, the complexity of the algorithm is polynomial to the number of users and available power levels (that depend on the size of the increment in the Power Update step).

In the original version of the algorithm of [8], an underestimation of the interference prices is likely to occur in some cases. This can be caused by problems in message exchange, for example, due to users' mobility or increased update time intervals for the interference prices, considering that updates are asynchronous for all users. The effect of this underestimation is the convergence of the algorithm to a nonoptimal solution. Moreover, as the number of user pairs increases, the highest allowable transmission power level is more likely to be chosen, since the previous problems escalate. This is not desirable, since it will often result in increased interference to a potentially large number of neighboring users, especially in the case that the interference is underestimated for the reasons mentioned above.

Therefore, in this work, a coefficient "α" is introduced in order to improve the scalability of the algorithm in case a large number of users are sharing the same spectrum band and to cope with uncertainties, such as large update intervals and problems in the message exchange mechanism. In both cases, there is a danger that the impact of the interference on other users due to the increase in transmission power

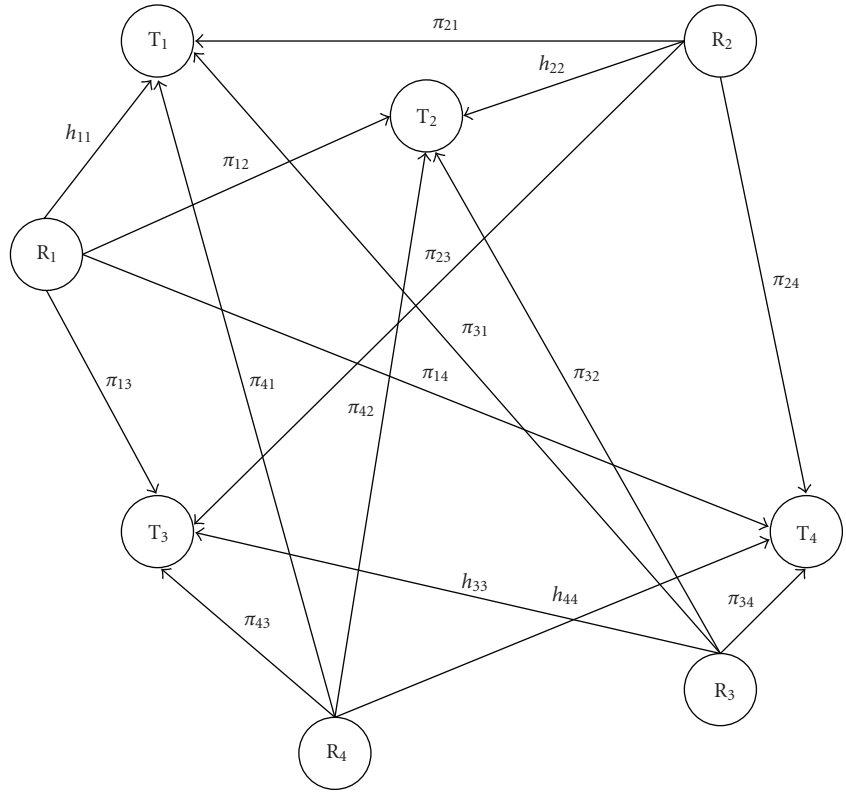


Figure 1: An example network topology with four transmitter-receiver pairs.

will be underestimated as explained above. Thus, factor α needs to avert this scenario by increasing the weight of the second term of (3), which expresses the utility loss other users will experience from a transmission power increase. In such cases, it will compensate for the underestimation of interference, by increasing the value of the second term and, therefore, it can result in a system that approximates the case of "perfect" message exchange (without long delays, reduced message range, etc. that reduce the second term in (3)).

If coefficient α is included as a weight multiplied with the subtracted interference term, then the following equation is derived, that is the objective to be maximized:

$$u_i\left(\gamma_i\left(p_i^k\right)\right) - a \cdot p_i^k \sum_{j \neq i} \pi_j^k \cdot h_{ji}. \tag{4}$$

In a "real" protocol implementation, parameters such as the storage requirements and scalability of the message exchange mechanism should be addressed. Moreover, the overhead and delays introduced by message exchange should be taken into consideration together with parameters such as timeliness and path optimality (for increased reliability in message transmission). However, the performance of the original version of the algorithm in [8] was shown not to degrade sharply in case the message exchange is imperfect (e.g., if the nodes can only exchange messages with their closest neighbors up to a specific range, or if some messages are lost). This characteristic is the outcome of the fact that, in the case of imperfect message exchange, the algorithm gracefully degrades towards the "worst case" scenario of unregulated transmission with the maximum allowable power level, as the value of the subtracted term is gradually underestimated in (3). The term "graceful degradation" refers to that fact that when a certain number of messages are lost, the performance of the system does not drop sharply towards the worst case. This characteristic is greatly desirable for systems that operate in faulty or unreliable environments (e.g., [10]). In this work, the previous property is further improved with the introduction of coefficient α that provides the capability to handle uncertainties.

3. Fuzzy Inference

Fuzzy logic is well suited for the purpose of defining the value of factor α since it can address vague and unclear requirements efficiently and the system can be easily fine-tuned to exhibit the desirable behavior. Fuzzy logic is based on fuzzy set theory, in which every object has a grade of membership in various sets. Inputs are mapped to membership functions or sets (fuzzification process). Knowledge of a restricted domain is captured in the form of linguistic rules. Relationships between two goals are defined using fuzzy inclusion and noninclusion between the support and hindering sets of the corresponding goals [11]. As a last step, the required output is defuzzified (to numerical) from the "THEN" part of the rules in order to produce the consequent.

An important advantage of fuzzy logic is that it can be applied transparently in combination with other well-known decision methods, such as multiobjective genetic algorithms [12] and game theoretic approaches [13]. Moreover, proper definition of the linguistic rules can be used to reduce signaling overhead by avoiding the ping-pong phenomenon, that is, when decisions or selections are made and the input variables are not constant but temporarily present regressive behavior. Network-related decision making and resource allocation based on fuzzy logic approaches have been proposed in various works, such as [14], with promising results.

For the previous reasons, but mainly due to its effectiveness in dealing with uncertainties and vague requirements, fuzzy logic was selected for defining the value of coefficient α, that is, the weight of the subtracted interference-related term in (4). Specifically, α is defined as:

$$\alpha = \frac{1}{\beta} \cdot \mathrm{IW} + \gamma, \tag{5}$$

Where IW is the Interference Weight derived after defuzzification. IW takes values in the range ($\beta_{\min}$–$\beta_{\max}$) in order to provide adequate resolution capabilities for the fuzzy reasoner, also according to the specific ranges of the membership functions. Parameter β has the value of $\beta_{\max}$, while γ equals 1. This implies that α cannot be greater than two, meaning that the underestimation of the interference is not expected to be greater than 100%. Beyond that point, message exchange is not considered very reliable and the algorithm degrades towards the "always transmit with the maximum power" case (although a portion of the underestimation is still alleviated). On the other hand, if uncertainties are very low, the first term of the sum is converging to zero and the value of the equation is approximately equal to that of the original algorithm. For all other cases the first term is a nonzero value in the (0,1) interval that compensates for a typical underestimation of the interference due to imperfect message exchange.

The fuzzy reasoner used for deriving α is of type "Mamdani", because it is intuitive, well suited for human input, flexible, and widely accepted. It receives three inputs (number of users, mobility level, and update time interval for the interference prices) and generates one output (the Interference Weight). The input membership functions are triangular (selected mainly for simplicity in calculations) and three membership functions per input variable are defined, therefore the number of fuzzy rules is $3^3 = 27$.

The membership functions for the output variable "Interference Weight" are five and the output value is set in the range (0–500), in order to achieve a greater degree of resolution and flexibility for the output of the fuzzy reasoner. The membership functions mf1–mf5 are given the labels "very low", "low", "moderate", "high" and "very high", respectively, in Table 1.

As can be seen, the number of users is selected to be the dominant factor, which has the greatest effect in the final outcome. This is a result of the fact that if the number of users is large, even a small increase in the transmission power of a user has the potential to cause increased interference and reduce the QoS to a large number of users if its effect is underestimated due to uncertainties in message exchange. The update time interval and the mobility level have similar weights but different behaviors. The first has a uniform effect over the entire valid range of update times; while the latter starts to affect the outcome only after a relatively high level, but after which it increases sharply, as only after a relatively high level of mobility is reached, users are likely to underestimate the interference they will cause to others (due to problems in message exchange, etc.).

Table 1: Rules of the fuzzy reasoner.

Rule number	Users	Update interval	Mobility level	Consequent
1	Low	Low	Low	Very low
2	Low	Low	Moderate	Very low
3	Low	Low	High	Low
4	Low	Moderate	Low	Low
5	Low	Moderate	Moderate	Low
6	Low	Moderate	High	Low
7	Low	High	Low	Low
8	Low	High	Moderate	Low
9	Low	High	High	Moderate
10	Moderate	Low	Low	Low
11	Moderate	Low	Moderate	Low
12	Moderate	Low	High	Moderate
13	Moderate	Moderate	Low	Moderate
14	Moderate	Moderate	Moderate	Moderate
15	Moderate	Moderate	High	High
16	Moderate	High	Low	Moderate
17	Moderate	High	Moderate	High
18	Moderate	High	High	High
19	High	Low	Low	Moderate
20	High	Low	Moderate	High
21	High	Low	High	High
22	High	Moderate	Low	High
23	High	Moderate	Moderate	High
24	High	Moderate	High	Very high
25	High	High	Low	High
26	High	High	Moderate	Very high
27	High	High	High	Very high

The Defuzzification method used for generating the final crisp value is "Centroid", also known as "Center of Gravity (COG)". This method determines the center of the area below the combined membership function; therefore the final output u_{COG} is given from (6), where u_i are the centers of the membership functions $\mu_F(u)$:

$$u_{\mathrm{COG}} = \frac{\sum_1^{27} u_i \cdot \mu_F(u_i) du}{\sum_1^{27} \mu_F(u_i) du}. \tag{6}$$

The defuzzification method takes into account the area as a whole, counting overlapping regions only once.

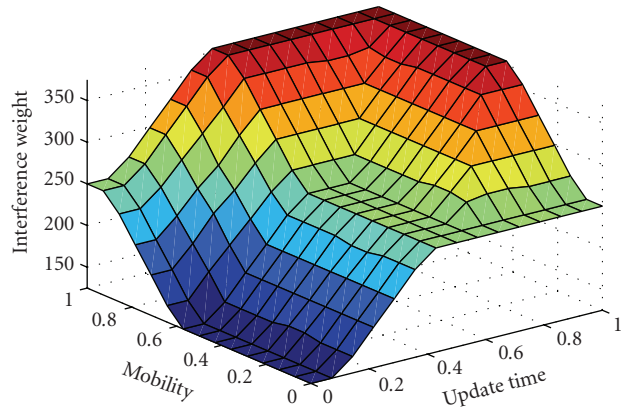


Figure 2: Interference weight as a function of the update interval and the mobility level.

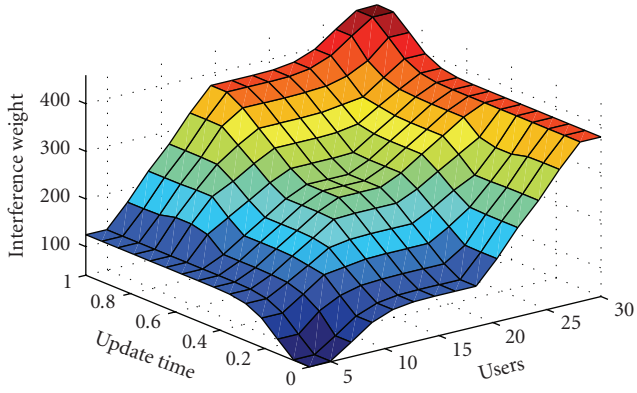


Figure 3: Interference weight as a function of the update interval and the number of users.

The three-dimensional (3D) representation of the Interference Weight (crisp value in the range (0–500)) as a function of the interference price update time interval and the mobility level is presented in Figure 2.

The coefficient increases with the update time interval as it is more likely that transmitters do not have the updated interference price information for other users. The increase is approximately uniform for the entire valid update time range. On the other hand, the coefficient also increases as the level of mobility increases. However, in this case the increase is not uniform but begins after a relatively high mobility level is reached and then rises quickly. The exhibited behavior is the outcome of the fuzzy rules defined in Table 1.

The 3D representation of the Interference Weight derived from the specified rule base and defuzzification method as a function of the interference price update time interval (defined as up to 100 seconds but normalized in the (0–1) range) and the number of users (up to 20 user pairs) is presented in Figure 3.

For the update time interval, the behavior is similar to that in the previous case. On the other hand, the coefficient also increases with the number of users. The increase is rather sharp (as determined by the rules in Table 1) and the value of the coefficient is rising quickly even for a relatively small

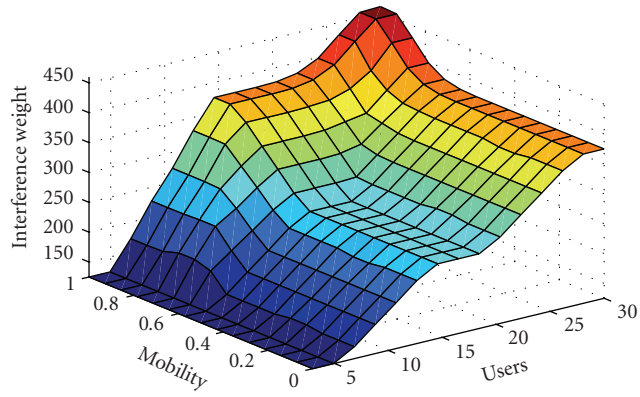


Figure 4: Interference weight as a function of the mobility level and the number of users.

number of users. This is necessary because, as mentioned previously, when the number of user pairs is large, even a small increase in interference has the potential to affect many users and significantly decrease the overall utility of the network.

The 3D representation of the Interference Weight as a function of the number of users and the mobility level, depicted in Figure 4, is presenting for both parameters the behavior explained above. The overall form of the figure resembles the previous; however the mobility level is starting to affect the outcome only after a threshold is crossed, as expected according to the selected set of fuzzy rules.

The overall methodology for the derivation of the optimal transmission power of every user pair is depicted in Figure 5. Initially, the number of user pairs is defined, together with the mobility level and the update time interval for the interference prices. As a next step, fuzzification of the values takes place in order to prepare them for elaboration in a fuzzy logic context. Following the fuzzification process, fuzzy reasoning based on a set of predefined rules (Table 1) is applied. These rules describe the desired behaviour of the system and define the impact of the input parameters (number of users, mobility level, and update time interval) in the value of the Interference Weight. After fuzzy reasoning is completed, the result is defuzzified to numerical, giving the crisp value of the Interference Weight. The topology characteristics are used to initialize the simulator and every user selects a valid initial value for the transmission power level p_i^k and the interference price π_i^k. Finally, the users proceed to update their transmission power levels and interference prices asynchronously in order to maximize (4). The process is completed when the system reaches a steady point in which no user is requesting to modify its transmission level.

4. Performance Evaluation

The performance of the proposed algorithm is evaluated through extensive MATLAB simulations. In this direction, the overall utility value of the algorithm is initially compared to the utility of a simple "always select the maximum

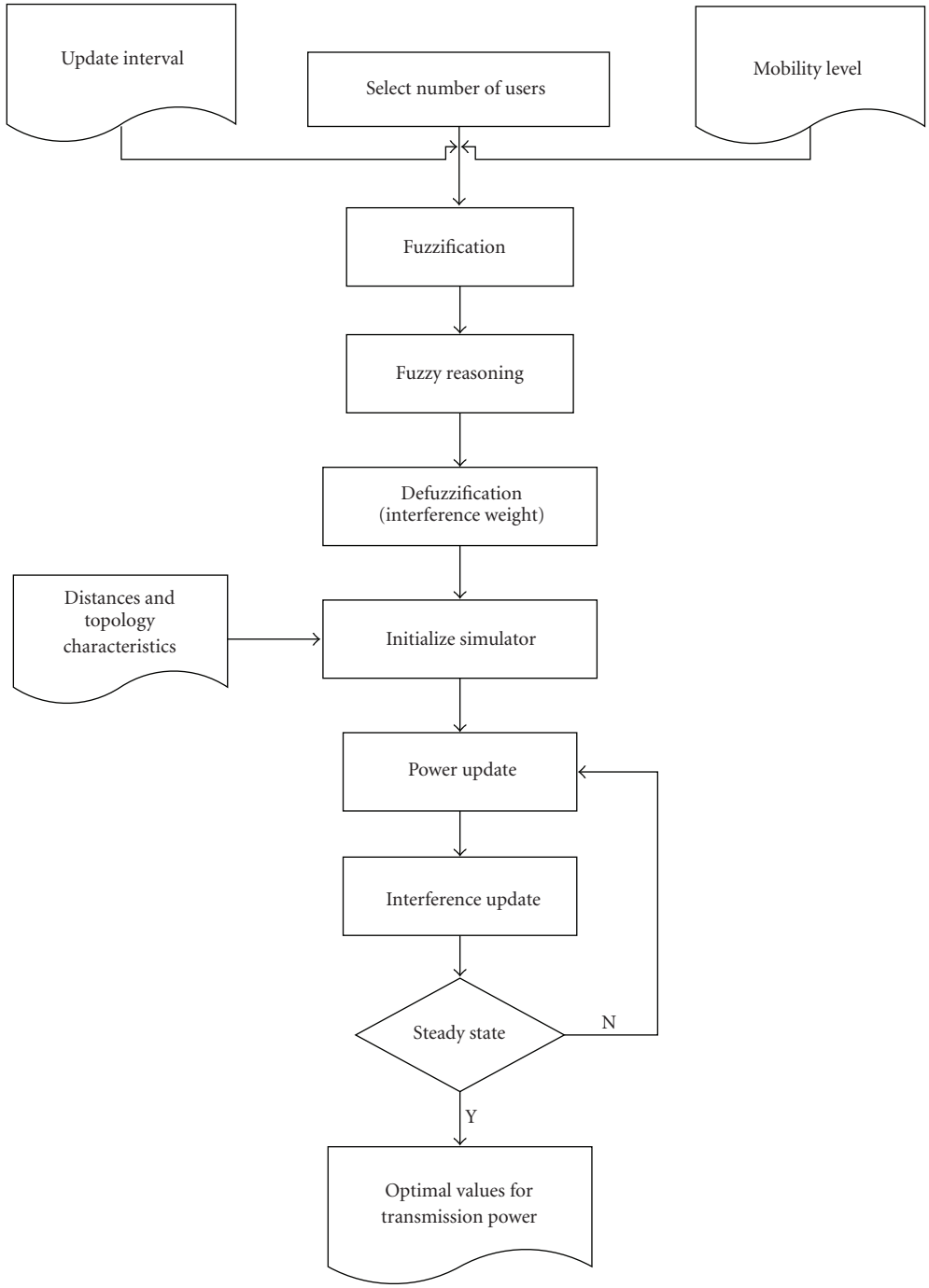


Figure 5: Overall methodology for deriving the transmission power levels.

valid power" policy as well as the utility of the original algorithm. The proposed algorithm is also applied in both FBMC and OFDM systems in order to validate its flexibility and capability of transparently exploiting an improved Physical layer, without any further modifications. Moreover, a scenario of long update time intervals in which some of the messages are delayed causing other nodes to not have the latest interference price information is considered, in order to study the performance of both algorithms in a specific case of nonideal message exchange. Finally, in order to quantify the improvement using conventional network metrics and to show the relation between a higher overall utility value and parameters that directly affect the user experience, comparison with the algorithm of [8] in terms of SINR is also performed.

As explained previously, users set their power level so as to maximize (4). The total "useful" utility for the network is the sum of the utilities for every user pair. The distance between users that constitute a pair is a random number in the (1–20) meters range, while the distance between users

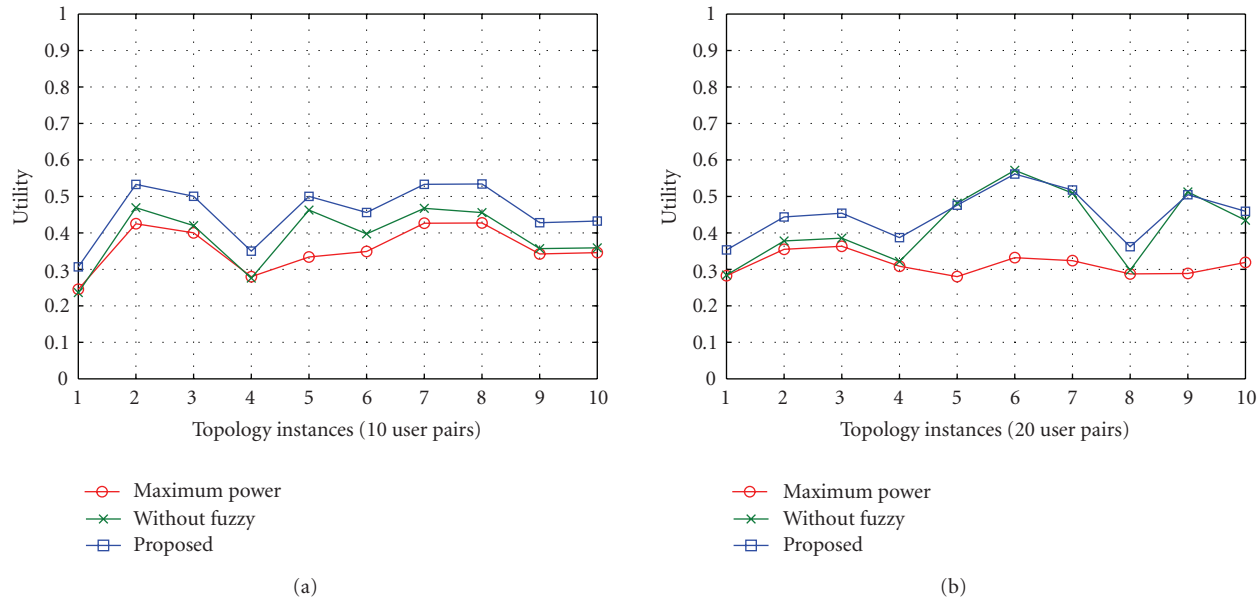


Figure 6: Utility values for the proposed algorithm, the always transmit with the maximum valid power scenario and the original algorithm without coefficient α for (a) 10 and (b) 20 user pairs.

that are not a pair is a random value in the $[5, 50]$ meters range. This is a more common and more practically significant scenario than using entirely random values, (e.g., it is often encountered in a conference room as well as an airport or train station, where coworkers are initiating a point-to-point ad hoc communication). The value of β in (5) is set to 500, since the Interference Weight takes values in the range (0–500), in order to provide adequate resolution capabilities. For all cases we assume the presence of uncertainties due to imperfect message exchange (one in every four messages is lost) that cause 25% underestimation of the interference. If such uncertainties are not present, then the algorithm behaves similarly to the algorithm of [8]. In the presence of uncertainties, parameter α compensates for the underestimation of interference and helps the system converge near its optimal point, as described in the previous sections.

The improvement in the total utility of the network if the proposed algorithm is utilized over the scenario in which every user transmits using the maximum allowable power level as well as over the original version of the algorithm that does not include the coefficient α, is depicted in Figure 6. The vertical axis depicts the achieved useful utility while the horizontal axis represents the corresponding topology instance. The considerable range over which the distance values are selected, coupled with the randomness of the relative positions between nodes and the presence of uncertainties that cause underestimation of the interference in ways that are not necessarily uniform (e.g., only some messages may be delayed), causes the final value of the utility function to vary significantly between different experiments both for the original and the proposed algorithm. Thus, the final utility of each topology instance is the average utility of ten experiments for the same instance. Finally, in order to study the effect of the number of users on the system, a scenario of 10 user pairs and 20 user pairs was simulated.

The utility for the scenario in which the users transmit always using the maximum power level defines the lower bound for the behaviour of the system. The proposed algorithm outperforms the original one, for the majority of times, with a more significant improvement for the lower utility values. This property is very important since it can improve the Bit Error Rate (BER) and raise QoS from poor to acceptable levels. Furthermore; the proposed algorithm always outperforms the always maximum power scenario, while the original algorithm in some cases results in similar performance. The reason for this is that the existence of the coefficient α in the proposed algorithm does not allow the system to reach the worst case of completely unregulated transmission since it always compensates for at least a portion of the underestimated interference. Another interesting point is that as the number of users increases, the average utility of the system decreases although extreme values are not affected significantly. This is justified by the fact that the interference exhibits a cumulative behavior that affects all other user pairs, therefore reducing the average utility. However, extreme values are mainly the outcome of the topology and the relative distance of the user pairs, thus, are less sensitive to the number of users.

The next step is to compare the results of the proposed algorithm using OFDM and FBMC systems. However, a short outline of the FBMC technique is required. According to the principle of transmission based on filter banks, the transmitter incorporates a Synthesis Filter Bank (SFB) while the receiver incorporates an Analysis Filter Bank (AFB). In the structure, the Fast Fourier Transform (FFT) is present as in OFDM [15]. It is however augmented, to complete a filter bank, by the Polyphase Network (PPN) which is comprised of a set of digital filters, whose coefficients globally form the impulse response of the so-called prototype low-pass filter. FBMC systems have somewhat increased hardware

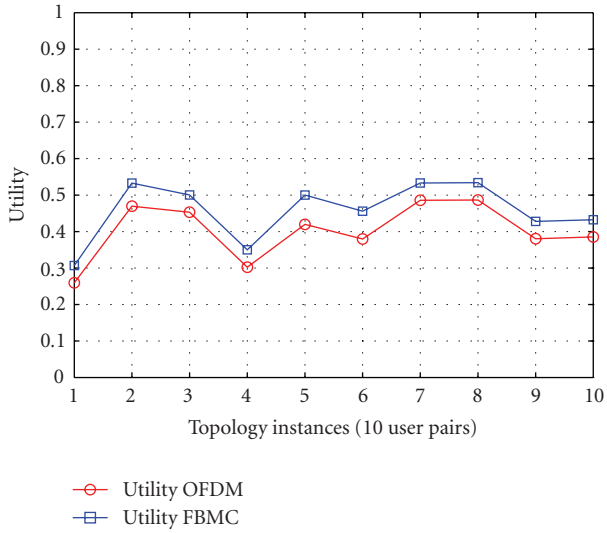


Figure 7: Utility function for the proposed algorithm with FBMC and OFDM.

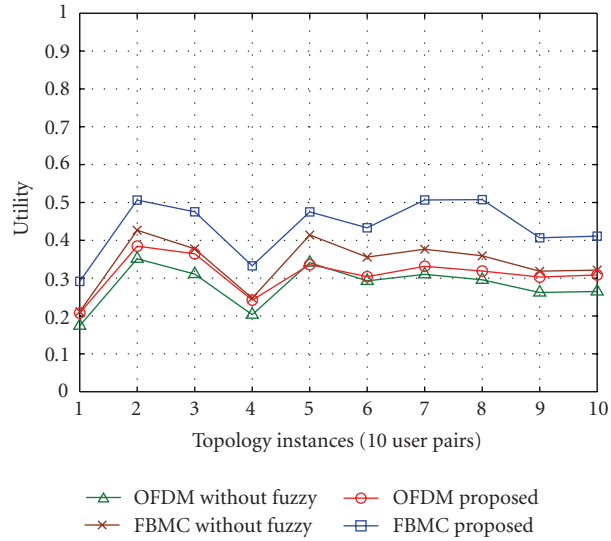


Figure 8: Utility values for the original and the proposed algorithm with both FBMC and OFDM, under the assumption of long message delays (10 user pairs).

complexity compared to the classical OFDM approach but compensate for this with a number of advantages. Among others, they do not require guard time and cycle prefix, while the use of Offset QAM (OQAM) implies that the full capacity of the transmission bandwidth is achieved. The improvement in the total utility of a network consisting of ten user pairs if the proposed algorithm is used with FBMC over OFDM is depicted in Figure 7.

This improvement stems from the fact that FBMC uses lower transmission power for the same bandwidth compared to OFDM [16] and therefore causes reduced interference. The proposed algorithm is able to transparently exploit this improvement and translate it in increased utility values.

To evaluate the resilience of the algorithm in the presence of long update time intervals we perform simulations with the assumption that some of the messages are delayed and, consequently, other nodes do not have the latest interference price information that has been announced. Thus, the definition of "long update times" that we consider in this work is to be at least equal to twice the average update time (so that other nodes have updated the announced interference price in this interval). Since it is already established that transmitting with the maximum power is the lower bound of performance for both the original and the proposed algorithm, in this scenario, we evaluate the behaviour of the original and the proposed algorithm with both FBMC and OFDM in order to study the effect of increased delays on each of these cases.

The first point that is noteworthy is the fact that the improved Physical layer of FBMC in this scenario provides a significant advantage that even surpasses the advantage offered by the proposed algorithm. Therefore, using FBMC with the original algorithm is better for this case than using OFDM with the proposed algorithm. The best option is to use the proposed algorithm with FBMC, combining the advantage of improved Physical layer capabilities and improved upper-layer functions. Regarding the latter point, the proposed algorithm consistently outperforms the original one when both use the same Physical layer (FBMC) under the assumption of long delays. Furthermore, if we juxtapose Figure 8 with Figure 6 we can derive some additional conclusions. Specifically, although the average utility values are reduced for all algorithms, the proposed algorithm is not affected as much as the original from the increased delays, thus the property of "graceful degradation" is indeed enhanced. Since real systems usually have to cope with nonideal conditions, this property is highly desirable.

Finally, it is very important to quantify the performance improvement in terms of conventional network metrics to show the relation between a higher overall utility value and parameters that directly affect the user experience. Since the main comparison is between the original algorithm and the proposed one, their behaviour in terms of SINR is compared in Figure 9. SINR is chosen as the most appropriate metric for comparison as it reflects directly on the QoS and the final user experience and can also be compared without considering external parameters, such as modulation and coding schemes that will impact for example the final BER of the system. The two graphs are following a similar pattern but the proposed algorithm consistently outperforms the original when the interference is underestimated, as it compensates for the interference underestimation and raises SINR to acceptable levels, especially for the lower values.

Regarding the overall simulation time and scalability properties of the algorithm, for all cases, the number of iterations for convergence is comparable to the number of user pairs. More specifically, for 10–30 user pairs usually less than thirty and up to fifty iterations are required for reaching the final steady state. Furthermore, the average execution time on a Core2 Quad Q9400 CPU operating at 2.66 GHz is less than two minutes for up to 20 user pairs and approximately five minutes for up to 30 user pairs.

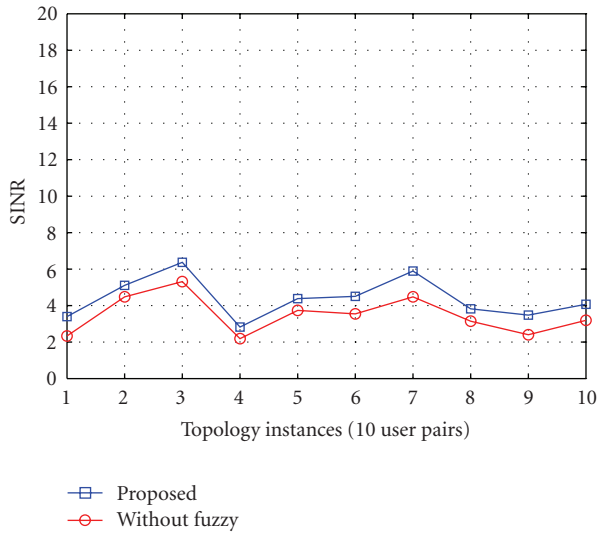


Figure 9: SINR for the original and the proposed algorithm under 25% underestimation of interference.

5. Conclusion

In this paper, an improved algorithm, based on the algorithm of [8], was presented for cooperative DSA in unlicensed bands, utilizing MAC layer mechanisms for message exchange (interference prices) between the secondary nodes in order to achieve interference mitigation. The main improvement in this work compared to [8] is the introduction of a coefficient α that is serving as the weight of the interference term, increasing its impact in cases of imperfect message exchange, long update time intervals for interference prices, as well as increased number of users. In such cases, the interference that is caused to other user pairs by an increase in the transmission power of a user is often underestimated, resulting in a convergence of the algorithm to a nonoptimal solution. In the presence of such uncertainties, if this underestimation is compensated by a properly defined weight parameter, the system approximates its optimal behavior as in the case of "perfect" message exchange.

The value of the weight parameter was derived from a fuzzy logic reasoner. Fuzzy logic was selected because it is particularly effective in dealing with uncertainties and vague requirements. Moreover, the outcome of the proposed algorithm has been compared to the original algorithm in terms of the overall utility level (defined as the sum of the user utilities) under uncertainties that cause 25% underestimation of interference. Furthermore, comparison was also made between the proposed algorithm in FBMC and OFDM systems. In this case, using FBMC increased the achieved utility. The improvement stems from the fact that FBMC uses lower transmission power for the same bandwidth compared to OFDM and therefore causes reduced interference. Additionally, a scenario of long update time intervals in which some of the messages are delayed causing other nodes to not have the latest interference price information was considered, and the performance of both algorithms in a case of nonideal message exchange was evaluated. Results indicate that the algorithm consistently outperforms previous schemes in terms of SINR under uncertainties and can transparently exploit the improved Physical layer offered by FBMC.

Acknowledgments

This work was performed in Project PHYDYAS which has received research funding from the Community's Seventh Framework program. This paper reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein. The contributions of colleagues from PHYDYAS consortium are hereby acknowledged.

References

[1] P. Pawelczak, S. Pollin, H.-S. W. So, A. Bahai, R. Prasad, and R. Hekmat, "Quality of service assessment of opportunistic spectrum access: a medium access control approach," *IEEE Wireless Communications*, vol. 15, no. 5, pp. 20–29, 2008.

[2] F. Wang, M. Krunz, and S. Cui, "Price-based spectrum management in cognitive radio networks," in *Proceedings of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '07)*, pp. 70–78, August 2007.

[3] L. Ma, X. Han, and C.-C. Shen, "Dynamic open spectrum sharing MAC protocol for wireless ad hoc networks," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 203–213, November 2005.

[4] L. Cao and H. Zheng, "Distributed spectrum allocation via local bargaining," in *Proceedings of the 2nd Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '05)*, pp. 475–486, September 2005.

[5] C. Peng, H. Zheng, and B. Y. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *Mobile Networks and Applications*, vol. 11, no. 4, pp. 555–576, 2006.

[6] J. Zhao, H. Zheng, and G.-H. Yang, "Distributed coordination in dynamic spectrum allocation networks," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 259–268, November 2005.

[7] M. Bloem, T. Alpcan, and T. Basar, "A stackelberg game for power control and channel allocation in cognitive radio networks," in *Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools*, Nantes, France, October 2007.

[8] J. Huang, R. A. Berry, and M. L. Honig, "Spectrum sharing with distributed interference compensation," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 88–93, November 2005.

[9] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, 2006.

[10] O. A. Rawashdeh, "Towards decentralized management of graceful degradation in distributed embedded systems," in

Proceedings of IEEE Dependable Systems and Networks Conference (DSN '08), Anchorage, Alaska, USA, June 2008.

[11] C. Carlsson and R. Fullér, "Fuzzy multiple criteria decision making: recent developments," *Fuzzy Sets and Systems*, vol. 78, no. 2, pp. 139–153, 1996.

[12] S. Buljore, M. Muck, P. Martigne, et al., "Introduction to IEEE P1900.4 activities," *IEICE Transactions on Communications*, vol. E91-B, no. 1, pp. 2–9, 2008.

[13] W. Wei and M.-J. Wang, "Fuzzy-MOGA-based traffic signal control at intersection," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 639–644, November 2003.

[14] A. Merentitis, E. Patouni, N. Alonistioti, and M. Doubrava, "To reconfigure or not to reconfigure: cognitive mechanisms for mobile devices decision making," in *Proceeding of the 68th IEEE Vehicular Technology Conference (VTC '08)*, pp. 1–5, September 2008.

[15] H. Zhang, D. Le Ruyet, and M. Terre, "Spectral efficiency comparison between OFDM/OQAM- and OFDM-based CR networks," *Wireless Communications and Mobile Computing*, vol. 9, no. 11, pp. 1487–1501, 2009.

[16] D. S. Waldhauser, L. G. Baltar, and J. A. Nossek, "Filter bank based multicarrier systems," in *Proceedings of Techniken, Algorithmen und Konzepte für Zukünftige COFDM Systeme (TakeOFDM '08)*, 2008.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 531857, 12 pages
doi:10.1155/2010/531857

Research Article

Nonconvex Optimization of Collaborative Multiband Spectrum Sensing for Cognitive Radios with Genetic Algorithms

Michele Sanna and Maurizio Murroni

Department of Electrical and Electronic Engineering, DIEE, University of Cagliari, 09123 Cagliari, Italy

Correspondence should be addressed to Michele Sanna, michele.sanna@diee.unica.it

Received 1 November 2009; Accepted 6 April 2010

Academic Editor: Marina Mondin

Cognitive Radio (CR) is a novel technology that permits secondary users (SUs) to transmit alongside primary users (PUs). PUs retain transparent communications whereas SUs perform spectrum sensing and adaptive transmission to avoid collisions. Ultra-wideband sensing is of primary importance for SU to sense and access opportunistically several bands at a time. Reliable detection in wide geographical regions needs collaborative sensing. Optimal collaborative multiband sensing is not analytically solvable unless some approximations and solution domain restrictions are applied for convexity exploitation. In this paper, we demonstrate that convex constraints are deleterious. We propose an alternative optimization technique based on genetic algorithms. Genetic programming performs a direct search of the optimal solution without approximations and solution domain restrictions. As a consequence, collaborative multiband sensing can be consistently optimized without limitations. Additionally the genetic optimization exploits the correlation of time-varying channels for fast adaptive convergence.

1. Introduction

Recent studies have revealed a deep underutilization of the electromagnetic resource due to the static allocation of spectrum band licenses [1]. Large portions of spectrum remain unexploited during certain periods of time and in certain geographical regions, yielding to an average 90% idleness of the licensed bands—the so-called *spectral holes*. The few unlicensed ISM bands are rapidly overloading due to the boom of wireless applications demanding for sporadic access to the spectrum. Cognitive Radio (CR) is a novel technology introduced with the intent of intelligently exploiting the unused spectrum [2, 3]. In a cognitive network, secondary users (SUs) are allowed to operate in licensed bands, under the condition that they do not interfere with the transmissions of the primary users (PUs), legal lessees of the license. Cognitive SUs detect the presence or the absence of transmissions, and identify the unused portions of spectrum. Then they may adapt their time-frequency transmission parameters in order to occupy the detected spectral holes. Some new standards are already adopting this paradigm. IEEE P.1900, for instance, is developing a whole type of ad-hoc wireless network based on Dynamic Spectrum Access (DSA) to the available spectrum [4]. IEEE 802.22 standard is considering the TV bands of the Ultra-High Frequencies (UHFs) for Wireless Regional Area Networks (WRANs) [5, 6].

Spectrum sensing is performed to detect the existing transmissions and identify the spectral holes. Reliable and extensive detection should be performed in order to avoid unwanted collisions and find as much free resources as possible. Observing an ultra-wide range of frequencies at a time is a challenging task since it requires expensive high-speed RF equipment. Although there have been some proposals of wavelet decomposition for multiresolution analysis [7], a common approach is to use tunable bandpass filters and observe one band at a time. Multiband Joint Detection proposed by Quan et al. in [8] applies narrowband sensing techniques to operate an opportunistic optimization of the throughput over multiple independent bands. The channel is divided into K nonoverlapping narrow subbands which may be utilized by distinct primary systems or may be blindly sensed, that is, without knowing who is transmitting there. Energy thresholds for each subband have to be chosen for optimal detection aimed at maximizing the throughput and limiting the interference.

Single CR sensing results can be affected by shadowing or multipath fading [9]. SU may cause harmful interference to the primaries for unreliable decision. Multiple geographically distributed CR can perform collaborative sensing by combining their results and improve performance. In Linear Statistics Combination (LSC), a simple fusion rule of the power levels is applied to perform a reliability-enhanced unique decision. The weights of a linear combination have to be determined by formulating an optimization problem. LSC has demonstrated to noticeably enhance the performance of individual sensors [10].

By applying LSC to multiband detection, Quan proposed the Spatial-Spectral Joint Detection [11]. Multiple independently-sensed levels are bandwise combined in order to increase the reliability in each subband. As a consequence, the useful throughput is increased while further limiting the interference.

Optimal multiband detection is achieved by formulating an optimization problem. Since in general the aforementioned formulations are nonconvex, some methods have been proposed to transform or approximate the problems and limit the solution domain in order to exploit the hidden convexity [8]. Such constraints bind the maximum interference and minimum channel utilization, with a loss of generality in the practical detection configurations.

In this paper we propose at first an alternative formulation of the multiband sensing and then an implementation of the genetic algorithms that solve the presented formulations of detection problems. Our formulation of the multiband sensing interprets the limited interference regime as a bound on the interference caused in each subband, rather than the aggregate disturbance throughout the wideband channel. This yields to a direct solution for the noncollaborative multiband problem. It also reduces the collaborative multiband sensing problem to a set of simple narrowband LSC optimizations, still keeping a controlled interference regime and the aggregate throughput as the objective of the maximization. We then propose the genetic programming as a solving technique that avoids reformulations, approximations, and limitations. Genetic Algorithms (GAs) are extensively used to find true or approximate solutions in different communications engineering applications such as, for instance, network design [12] and adaptive modulations [13]. GAs perform a direct search of the best solution by considering sets of candidates and evaluating them singularly by means of a fitness measure, objective of the maximization. Then they iteratively drop unfit elements and select the fittest ones for combination (reproduction) and random alteration (genetic mutation) [14]. GAs work one step above mathematical analysis, dealing well with nondifferentiable functions as well as functions with multiple local maxima. Exploitation of hidden convexity is not necessary so that no reformulations or approximations of the problem are performed. We show how the genetic programming, by working directly at the root of the problem, can be an efficient and powerful optimization and analysis tool for all possible CR systems. Additionally, GAs demonstrate to exploit the correlation of different detection conditions with time-varying channels. Moving CR senses a channel with variable statistics. GAs exploit the statistical dependence of consecutive sensing events by performing an optimization that starts from the result of the previous instant. The sensing precision is drastically improved as well as the convergence time.

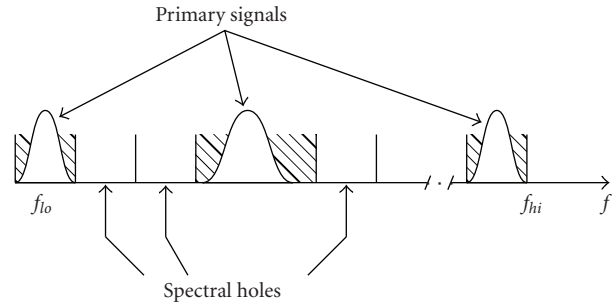


Figure 1: Schematic representation of the multiband channel.

The paper is organized as follows. In Section 2, the multiband detection framework is presented. Section 3 introduces the collaborative sensing within the opportunistic multiband fashion. Section 4 analyzes the genetic algorithms and the possibilities of application. In Section 5 are produced the generic numerical results, whereas Section 6 introduces the results for adaptive optimization and Section 7 concludes the paper.

2. Opportunistic Multiband Sensing

Multiband sensing considers a channel divided into K narrow subbands where one or more primary communication systems may be transmitting. A cognitive SU applies narrowband sensing to each subchannel in order to maximize its own transmission without harm for the PU.

2.1. Narrowband Signal Detection. A CR senses constantly the spectrum to discover which of the K subchannels are free of primary transmissions (Figure 1). Deciding for the condition of the single kth subband means posing the following binary condition [15]:

$$\begin{aligned} &\mathcal{H}_{0,k} : X_k = V_k, \\ &\mathcal{H}_{1,k} : X_k = H_k S_k + V_k, \quad k = 1, \ldots, K, \end{aligned} \tag{1}$$

where $\mathcal{H}_0$ represents the absence of primary signal (only Gaussian noise V_k with power σ_v^2) and $\mathcal{H}_1$ represents the presence of primary signal S_k, corrupted by Gaussian noise V_k. Capital letters indicate that we are considering the frequency spectra. H_k is the channel gain between the primary transmitter and the secondary receiver. The presence or absence of the primary signal in each subband is then verified as follows:

$$Y_k \triangleq \sum_{n=1}^{N} |X_k(n)|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_k, \quad k = 1, \ldots, K, \tag{2}$$

where γ_k is the subband energy threshold. The test statistics Y_k can be considered asymptotically normally distributed if

N is large enough (e.g., $N > 10$). The performance of the detection is calculated in terms of

(i) probability of identifying the spectral hole:

$$P(\mathcal{H}_{0,k} \mid \mathcal{H}_{0,k}) = 1 - P(\mathcal{H}_{1,k} \mid \mathcal{H}_{0,k}) = 1 - P_f^{(k)}, \quad (3)$$

(ii) probability of missed detection (interference):

$$P(\mathcal{H}_{0,k} \mid \mathcal{H}_{1,k}) = 1 - P(\mathcal{H}_{1,k} \mid \mathcal{H}_{1,k}) = 1 - P_d^{(k)}, \quad (4)$$

where $P_f^{(k)}$ is the probability of false alarm and $P_d^{(k)}$ is the probability of detecting a primary signal. The random variables Y_k have Gaussian distribution $Y_k \sim \mathcal{N}(\mu_{0,k}, \sigma_{0,k}^2)$ under hypothesis $\mathcal{H}_{0,k}$ and $Y_k \sim \mathcal{N}(\mu_{1,k}, \sigma_{1,k}^2)$ under $\mathcal{H}_{1,k}$.

Thus, $P_f^{(k)}$ and $P_d^{(k)}$ can be calculated as the tail of a normal distribution:

$$P_f^{(k)} = P(\mathcal{H}_{1,k} \mid \mathcal{H}_{0,k}) = Q\left(\frac{\gamma_k - \mu_{0,k}}{\sigma_{0,k}}\right), \quad (5)$$

$$P_d^{(k)} = P(\mathcal{H}_{1,k} \mid \mathcal{H}_{1,k}) = Q\left(\frac{\gamma_k - \mu_{1,k}}{\sigma_{1,k}}\right). \quad (6)$$

Increasing the utilization of the channel implies higher interference to the PU. The thresholds have to be optimized in order to maximize the utilization while limiting the interference.

2.2. Multiband Detection. Opportunistic multiband detection optimizes the aggregate throughput throughout the K subbands while limiting the interference (Figure 1). The design objective is to find an optimal vector of thresholds $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_K]^T$ such that the free spectrum holes are efficiently exploited and a controlled level of interference is produced. The probabilities of false alarm and detection can be represented compactly as follows:

$$\begin{aligned} \mathbf{P}_f(\boldsymbol{\gamma}) &= \left[P_f^{(1)}(\gamma_1), P_f^{(2)}(\gamma_2), \ldots, P_f^{(K)}(\gamma_K)\right]^T, \\ \mathbf{P}_d(\boldsymbol{\gamma}) &= \left[P_d^{(1)}(\gamma_1), P_d^{(2)}(\gamma_2), \ldots, P_d^{(K)}(\gamma_K)\right]^T. \end{aligned} \quad (7)$$

Given that $\mathbf{1} - \mathbf{P}_f(\boldsymbol{\gamma})$ denotes the probabilities of detecting the free subbands, r_k denotes the achievable throughput over the kth subband and that $\mathbf{r} = [r_1, r_2, \ldots, r_K]^T$, we can express the aggregate opportunistic throughput reached over the K subbands as

$$R(\boldsymbol{\gamma}) = \mathbf{r}^T\left[\mathbf{1} - \mathbf{P}_f(\boldsymbol{\gamma})\right]. \quad (8)$$

Similarly, if $\mathbf{1} - \mathbf{P}_d(\boldsymbol{\gamma})$ is the vector of probabilities of interference and $\mathbf{c} = [c_1, c_2, \ldots, c_K]^T$ is a measure of cost caused by transmitting in the subbands, then the aggregate interference can be expressed as

$$I(\boldsymbol{\gamma}) = \mathbf{c}^T[\mathbf{1} - \mathbf{P}_d(\boldsymbol{\gamma})]. \quad (9)$$

The maximization of (8) and the minimization of (9) are conflicting tasks. According to the formulation suggested by Quan et al. in [8], we limit the per-band interference ($1 - P_d^{(k)} \leq \alpha_k$), as well as the aggregate interference (9), and ensure a minimum utilization ($1 - P_f^{(k)} \geq \beta_k$). The optimization becomes finding the appropriate thresholds $\boldsymbol{\gamma}$ that maximize (8) with the aforementioned bounds [8]:

$$\begin{aligned} \max_{\boldsymbol{\gamma}} \quad & R(\boldsymbol{\gamma}) \\ \text{s.t.} \quad & I(\boldsymbol{\gamma}) \leq \epsilon, \\ & \mathbf{1} - \mathbf{P}_d(\boldsymbol{\gamma}) \leq \boldsymbol{\alpha}, \\ & \mathbf{1} - \mathbf{P}_f(\boldsymbol{\gamma}) \geq \boldsymbol{\beta}. \end{aligned} \quad (\text{P1})$$

The subchannel interference and utilization bounds can be translated to the linear constraint

$$\gamma_{\min,k} \leq \gamma_k \leq \gamma_{\max,k}, \quad k = 1, 2, \ldots, K \quad (10)$$

to be imposed to (P1), where

$$\begin{aligned} \gamma_{\min,k} &= \mu_{0,k} + \sigma_{0,k} Q^{-1}(1 - \beta_k), \\ \gamma_{\max,k} &= \mu_{1,k} + \sigma_{1,k} Q^{-1}(1 - \alpha_k). \end{aligned} \quad (11)$$

Problem (P1) is convex if the utilization is at least 50% ($\beta_k \geq 0.5$) and the interference at most 50% ($\alpha_k \leq 0.5$). Although these prerequisites are reasonable in many cases, convex maximization is not able to solve (P1) without these impositions.

In Section 4, we introduce the the genetic algorithms to solve the optimization for nonconvex CR systems, including the cases $\beta_k < 0.5$ and $\alpha_k > 0.5$.

3. Collaborative Multiband Sensing

The introduction of cooperative sensing aims at improving the sensing reliability in order to reduce the interference. Let us consider now a set of M CR operating in the primary region (Figure 2). The sensors detect individually the presence of primary transmissions by sensing the surrounding channel in each narrow subband and pose the binary condition (1).

3.1. Reliable Sensing. The receiving conditions of one single CR are subject to fading so that the sensing of only one single terminal can produce unreliable results (Figure 2) [9].

(i) Hidden terminal: a sensor (CR1) is located behind an obstacle. It senses a low power signal from PTx and may thus decide for $\mathcal{H}_0$ also when the transmission is present, affecting the reception of PRx.

(ii) Far terminal: a sensor (CR2) lies outside the primary range. It receives a low power level due to the distance and thus decides for $\mathcal{H}_0$. Its transmission can produce interference to PRx, which is inside the primary range.

Space diversity is exploited in CS to increase the reliability of the decision by combining the sensed power levels in a spatially distributed channel.

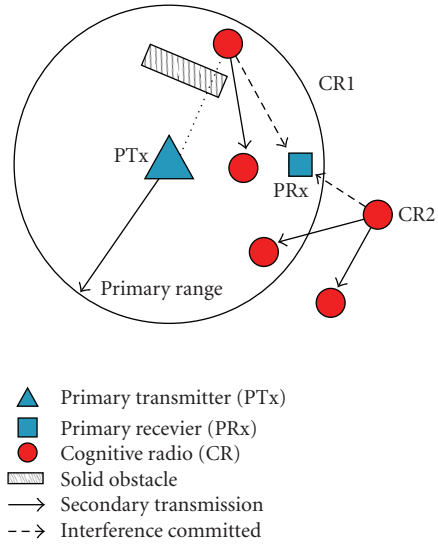


Figure 2: Examples of secondary interferences by cognitive sensors.

3.2. Collaborative Detection. Let us consider the signal received by each CR in one single narrow subband. The structure of a narrowband energy detector is (2) for each kth subchannel.

The sensing results are collected channelwise into a vector $\mathbf{Y}_k \triangleq [Y_k(1), Y_k(2), \ldots, Y_k(M)]^T$. Then, for each band, the M sensed levels are linearly combined with a vector of weights $\mathbf{w}_k$ and compared to the subband threshold γ_k [10]:

$$\mathbf{w}_k^T \mathbf{Y}_k = \sum_{i=1}^{M} w_k(i) Y_k(i) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_k, \quad k = 1, \ldots, K. \tag{12}$$

The weight factors $0 \le w_k(i) \le 1$ represent the contributions of the different CR in the respective subchannels. A high weight is more likely to correspond to a sensor with good SNR. The probabilities of false alarm (5) and detection (6) are now expressed as:

$$P_f^{(k)}(\mathbf{w}_k, \gamma_k) = Q\left(\frac{\gamma_k - \mathbf{w}_k^T \boldsymbol{\mu}_{0,k}}{\sqrt{\mathbf{w}_k^T \boldsymbol{\Sigma}_{0,k} \mathbf{w}_k}}\right), \tag{13}$$

$$P_d^{(k)}(\mathbf{w}_k, \gamma_k) = Q\left(\frac{\gamma_k - \mathbf{w}_k^T \boldsymbol{\mu}_{1,k}}{\sqrt{\mathbf{w}_k^T \boldsymbol{\Sigma}_{1,k} \mathbf{w}_k}}\right). \tag{14}$$

Throughput and interference depend now on the weight matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_K]$ and the threshold vector $\boldsymbol{\gamma}$:

$$R(\mathbf{W}, \boldsymbol{\gamma}) = \mathbf{r}^T \left[\mathbf{1} - \mathbf{P}_f(\mathbf{W}, \boldsymbol{\gamma})\right], \tag{15}$$

$$I(\mathbf{W}, \boldsymbol{\gamma}) = \mathbf{c}^T [\mathbf{1} - \mathbf{P}_d(\mathbf{W}, \boldsymbol{\gamma})]. \tag{16}$$

Coherently with the noncollaborative formulation, the collaborative multiband detection as proposed by Quan et al. in [8] is

$$\begin{aligned} \max_{\mathbf{W}, \boldsymbol{\gamma}} \quad & R(\mathbf{W}, \boldsymbol{\gamma}) \\ \text{s.t.} \quad & I(\mathbf{W}, \boldsymbol{\gamma}) \le \epsilon, \\ & \mathbf{1} - \mathbf{P}_d(\mathbf{W}, \boldsymbol{\gamma}) \le \boldsymbol{\alpha}, \\ & \mathbf{1} - \mathbf{P}_f(\mathbf{W}, \boldsymbol{\gamma}) \ge \boldsymbol{\beta}. \end{aligned} \tag{P2}$$

The previous considerations apply. This formulation provides further complications for convex optimization. The maximization is not convex and has to be lower bounded with another convex function. This brings loss of performance in terms of achievable throughput [8].

In the next section we introduce an alternative formulation of the multiband detection problem that can be solved by maximizing the LSC.

Then in Section 4 we introduce GA to solve directly the joint detection problem without the limitation of the convexity constraints.

3.3. Multiband Detection without Aggregate Constraint: LSC Optimization. Multiband detection without aggregate constraint can be seen as a collaborative detection with multiband aggregation. This alternative formulation of the multiband detection problem aims at achieving performance optimization with linear collaboration maximization applied to the single bands, still maximizing the aggregate throughput. LSC problem is less complex compared to the aforementioned formulations, which results in a faster and more precise solution with GA.

Let the secondary CR sense K subchannels which are licensed each one to a different primary system. Then the aggregate interference (16) caused by the SU is not a relevant index of caused harm because it is distributed between distinct PUs. It is rather more critical to limit the interference in the individual bands $P_m^{(k)}$ and look for the best weights and thresholds that maximize the throughput (15):

$$\begin{aligned} \max_{\mathbf{W}, \boldsymbol{\gamma}} \quad & R(\mathbf{W}, \boldsymbol{\gamma}) \\ \text{s.t.} \quad & \mathbf{P}_m(\mathbf{W}, \boldsymbol{\gamma}) = \mathbf{1} - \mathbf{P}_d(\mathbf{W}, \boldsymbol{\gamma}) \le \boldsymbol{\alpha}. \end{aligned} \tag{P3}$$

No minimum utilization is imposed, given that it may be not always necessary. A weight optimization for the linear statistic combination must be performed in each subband, in order to maximize the subchannel utilization:

$$\begin{aligned} \max_{\gamma, \mathbf{w}_k} \quad & P(\mathcal{H}_{0,k} \mid \mathcal{H}_{0,k}) \\ \text{s.t.} \quad & P(\mathcal{H}_{0,k} \mid \mathcal{H}_{1,k}) \le \alpha_k, \end{aligned} \qquad k = 1, \ldots, K. \tag{P3a}$$

We can explicit the threshold from the interference constraint by means of (14)

$$\gamma_k = Q^{-1}(1 - \alpha_k) \sqrt{\mathbf{w}_k^T \boldsymbol{\Sigma}_{1,k} \mathbf{w}_k} + \boldsymbol{\mu}_{1,k}^T \mathbf{w}_k, \tag{17}$$

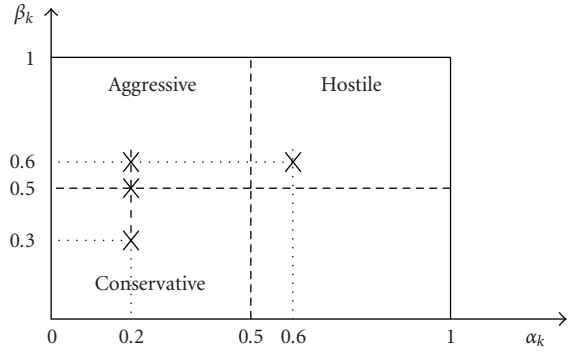


Figure 3: Schema of the CR classes plane. Case studies of Figure 5 are indicated with a cross.

and we can solve for $k = 1, \ldots, K$, the following unconstrained maximization:

$$\max_{\mathbf{w}_k} \frac{Q^{-1}(1-\alpha_k)\sqrt{\mathbf{w}_k^T \boldsymbol{\Sigma}_{1,k} \mathbf{w}_k} + \left(\boldsymbol{\mu}_{1,k} - \boldsymbol{\mu}_{0,k}\right)^T \mathbf{w}_k}{\sqrt{\mathbf{w}_k^T \boldsymbol{\Sigma}_{0,k} \mathbf{w}_k}}, \quad (18)$$

where $\boldsymbol{\mu}_{0,k}$ and $\boldsymbol{\Sigma}_{0,k}$ are the vector of the M average power levels and the matrix of covariances relative to the kth subband in case of $\mathcal{H}_0$, respectively—$\boldsymbol{\mu}_{1,k}$ and $\boldsymbol{\Sigma}_{1,k}$ in case of $\mathcal{H}_1$, respectively.

Function (18) is again nonconvex in general. Convex maximization is feasible in the case $P_d \leq 0.5$, where hidden convexity is exploitable. Then separate convex subdomains may be considered for the application of techniques such as Semidefinite Programming (SDP), which requires complex reformulations in matrix forms [16]. The eventuality of a system with nonexploitable convexity is not far from reality especially when few CRs collaborate together. Other solving techniques have to be implemented.

3.4. CR Systems Classification. A common classification of CR systems is based on the per-band utilization and interference (Figure 3) [10]

(i) *Conservative System*: A *conservative* CR system has a utilization less or equal to 50% ($\beta_k \leq 0.5$) and a probability of interference smaller than 50% ($\alpha_k < 0.5$).

(ii) *Aggressive System*: An *aggressive* CR system achieves a utilization higher than 50% ($\beta_k > 0.5$) and an interference less than 50% ($\alpha_k < 0.5$).

(iii) *Hostile System*: A *hostile* CR system targets more than 50% of spectrum utilization ($\beta_k > 0.5$), causing at least 50% of interference ($\alpha_k \geq 0.5$).

The Q(x) function is convex if $x > 0$ and concave if $x \leq 0$. Consequently, a necessary condition for (P1) and (P2) to have convex objective with concave constraints is that

$$\beta_k \geq 0.5, \quad \alpha_k \leq 0.5, \quad k = 1, \ldots, K, \quad (19)$$

that is, the system must be *aggressive*. For the problem (P1) this is also a sufficient condition, whereas when considering the weights (P2) the objective throughput is to be additionally lower bounded by a function, for which the aforementioned conditions are sufficient for convexity. The lower bound brings a loss of performance as documented in [8]. The convexity region includes the values from the *conservative* region where $\beta_k = 0.5$ and from the *hostile* systems where $\alpha_k = 0.5$ (Figure 3). The inside of the other regions is in any case not solvable with convex maximization.

The α_k and β_k values are bounds for the minimum and maximum per-band threshold by means of (11), which limit above and below aggregate interference and throughput. An exemplary behavior is that an *aggressive* system may not transmit with low bitrate because a minimum per-band utilization is imposed. It can not either cause less disturbance than a lower implied threshold by the utilization bound.

Section 4 introduces thus the genetic algorithm for solving without convexity limitations these three sensing problems.

(i) Individual Multiband Detection (P1).

(ii) Collaborative Multiband Detection (P2).

(iii) Collaborative Multiband Detection without aggregate interference constraint—LSC (P3).

We show how the genetic approach performs a direct search of the solution that provides the highest throughput by generating, comparing, and discarding various solutions. The advantages of such approach are that it does not need any problem reformulations or mathematical constraints. It is thus suggested as an acceptable way of solving directly the sensing problems for all CR system classes.

4. Optimization Using Genetics

Genetic Algorithms (GA) belong to the class of the evolutionary models for solving optimization and maximization problems (OP) [14]. The natural evolution processes are simulated by means of *natural selection* and *survival of the fittest* in order to find the maximum of a function. GA demonstrated optimum performance with complex multidimensional OP. Nonconvex maximizations (P1), (P2), and (P3) are problems where a lot of methods fail or manage to solve only a subset of cases.

4.1. The Genetic Algorithms. GAs consider a *population* of potential solution vectors to the OP and iteratively select good elements and drop unfits to let the best ones survive from generation to generation. The independent variables are called *genes* and they form a vector called *chromosome* $\mathbf{g}_k^{(n)}$. The numerical realization of such variables (the *genotype*) is a potential solution to the problem—also referred to as an *individuals* or an *element* of a population. The genes are different for the noncollaborative and collaborative problems. Namely the chromosome is

$$\mathbf{g}_k^{(n)} = \begin{cases} \boldsymbol{\gamma}_k^{(n)}, & \text{for (P1)}, \\ \left[\mathbf{w}_{1,k}^{(n)}, \mathbf{w}_{2,k}^{(n)}, \ldots, \mathbf{w}_{M,k}^{(n)}, \boldsymbol{\gamma}_k^{(n)}\right] & \text{for (P2)}, \\ \mathbf{w}_k^{(n)}, & \text{for (P3)}. \end{cases} \quad (20)$$

The genes are the independent variables of the throughput function (8), which is the objective of the maximization. A generation is a set of genotypes at the nth iteration. The subscript $k = 1, \ldots, \text{Pop}_{\text{size}}$ distinguishes the genotypes of a certain generation:

$$G_n = \left\{ \mathbf{g}_1^{(n)}, \mathbf{g}_2^{(n)}, \ldots, \mathbf{g}_{\text{Pop\_size}}^{(n)} \right\}. \tag{21}$$

The index of the generation is $n = 0, 1, 2, \ldots, N_{\text{gen}}$. In order to determine which genotypes fit the problem, a *fitness score* is calculated for each element:

$$\mathbf{s}(n) = \left\{ f\left(\mathbf{g}_1^{(n)}\right), f\left(\mathbf{g}_2^{(n)}\right), \ldots, f\left(\mathbf{g}_{\text{Pop\_size}}^{(n)}\right) \right\}. \tag{22}$$

The fitness function is the objective of the maximization, that is, the throughput (15) in (P1) and (P2) and (18) in (P3). Consecutive generations of solutions and their scores are examined iteratively in order to determine which individuals are suitable for surviving to the next generation as shown in Figure 4.

The generation zero is created randomly, whereas each generation $n + 1$ is created recursively in three steps.

(1) *Selection*: An intermediate population is created by performing the (natural) selection. The individuals with the best fitness scores are duplicated (in a predefined percentage) and the rest are discarded.

(2) *Crossover*: The intermediate population is recombined to simulate the reproduction. Survived elements are taken couplewise and, according to a mixing criterion, a number of children members are created from each couple.

(3) *Mutation*: A percentage of the offspring randomly mutates to create new genotypes. Spreading the search at each generation avoids restraining in a local maximum due to deception (see Section 4.2).

After these steps a new generation G_{n+1} is created from the previous one, with closer elements to the optimal solution (Figure 4). The children elements that do not respect the constraints are dropped. An equal number of elements is regenerated and the constraint is reevaluated, which can increase the computational load.

The evaluation and generation steps are performed iteratively, in order to increase the percentage of fit members. The computation is terminated when the fitness of the population remains unchanged for a sufficient lapse of generations or if the maximum number of generations Max _gen is reached. The genotype with the highest fitness is chosen as the final solution. The convergence precision depends on multiple factors and influences directly the time consumption of the process, as it is discussed in the following Section. For more details about GA refer to [14, 17].

4.2. Feasibility of the Genetic Approach. GAs are introduced because of the limitations observed when treating the intrinsically nonconvex detection problem with convex optimization methods. There is a single convex domain in (P1) and (P2), with the constraints (19) [8]. LSC (18) has two convex subdomains under a maximum interference bound, to be treated separately. Each one can be optimized with complex methods such as the SDP technique [16]. Other methods like the so-called *hill-climbing* methods need the objective function to be well behaved, that is, it has to be continuous as well as its derivatives. The function also has to be *unimodal*, that is, with only one peak, because with many peaks the search may stop at the first undergone relative maximum.

The main motivation for using GA is that they solve such complex multidimensional problems without in-depth function study, constraints, or reformulations. GAs do not have mathematical limitations such as the convexity requirement. They abstract from the smoothness of the objective function, because they calculate isolated points ignoring discontinuities, cusps, and inflections. GAs also perform well in presence of multiple relative maxima (e.g., in presence of ripples) by spreading the population variety and evaluate as much genotype variety as possible. This way of working one step above the complications of function analysis makes GA suitable for solving the multiband detection problem with any values of α_k and β_k.

Drawbacks of genetic programming are the *deception* and the computational load. The *deception* is the surviving of an apparently fit subpopulation that leads away from the global optimum [18]. This is equivalent to say that local maxima may cause ambiguity that let a GA converge away from the global maximum or not converge at all in reasonable time (slow finishing). Although this is in general not desirable, it has been proven that it is worth using GA if the OP to be solved presents a certain degree of deception, whereas regular-behaved problems are better solved with other methods [18]. In fact, GA have a strong attitude at escaping from local maxima by spreading around the search. Setting up appropriately the parameters of our GA is fundamental to find an optimal configuration for each kind of problem.

GAs also have a characteristic that makes them profitable in time-varying channels, such as the mobile radio channel. The channel statistics $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, i = 0, 1\}$ vary due to the movement of the sensors with respect to the PTx. At each sensing instant, reliable cooperation requires an optimization of the weights to adapt to the new statistics. Performing a new optimization procedure at each sensing instant is inefficient, if it is done with SDP or other convex maximization methods. Since the statistics are supposed not to vary too much, the weights of each sensing instant are correlated with the preceding ones. GAs can keep memory of the previous weights and use them as the starting point for the new elaboration. The convergence is faster since the starting vector should be already close to the new optimum.

4.3. Computational Cost. The computational effort is measured as the number of function evaluations that have to be done to complete a computation. As function evaluations we consider both fitness evaluations and constraints verifications indistinctly, since they have the same form and imply the same number of floating point operations. Comparisons

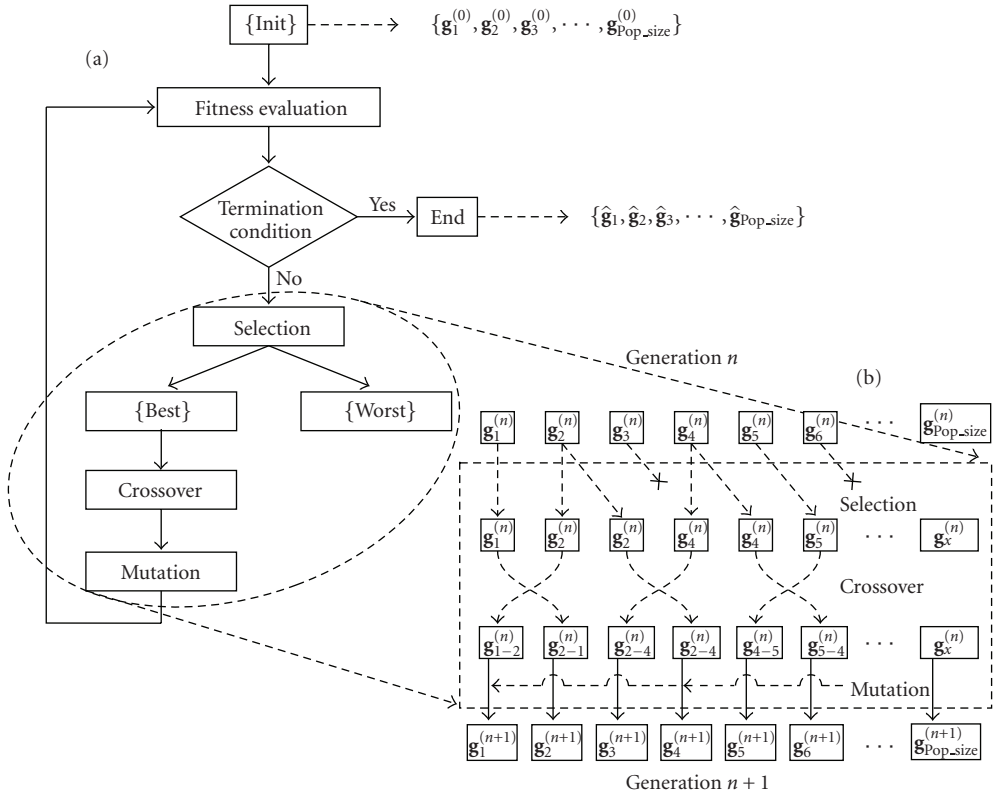


Figure 4: Flowchart of the proposed GA (a) and the three steps for the creation of a generation (b).

and data duplications are negligible. In order to converge with a certain accuracy our GA may need big populations and/or more generations, whereas small populations may result in insufficient precision or slow finishing.

The number of computations grows linearly with the population size. Since the aggregate interference constraint cannot be explicited, it has to be computed for each member. Those members that do not respect the constraint are regenerated and a further constraint evaluation is performed. In (P1) the verification of the per-band interference and per-band utilization (10) is made in the genotype space, so they do not yield to other function evaluations and no further increase of the computational load. All constraints in (P2) have instead to be verified for each generation. In general if we name h the total number of function evaluations that have to be computed for each member, then the total number of computations for a genetic optimization is

$$N_{\text{eval}} = h * \text{Pop}_{\text{size}} * N_{\text{gen}}. \tag{23}$$

In unconstrained optimization such as (P3), $h = 1$. In multiband problems (P1) and (P2), $h \geq 2$, because of one fitness and one constraint verification. The value grows as much as how many regenerations are needed in average to find one constraints-fit member. This depends on the crossover criterion, on the random mutation and on the stochastic realizations of the process, as well as on the problem itself. This index of computational effort is quite critical because before establishing a valid offspring the algorithm can undergo several computations.

5. Simulation Results

We analyze now the performance achieved by a multiband collaborative detection framework. The optimization is conducted with variable bounds in order to compare convexity-limited and nonconvex systems. The testbed channel has $K = 8$ subbands, secondary rates between 400 and 1000 kbps, and individual interference costs between 0.8 and 8. The SNR in each band is between -6 and -1 dB. The length of the detection interval is $N = 100$. Throughput-interference characteristics (I-R) are depicted for the three CR classes in Figure 5. The collaborative and noncollaborative systems differ for the LSC in the subbands. Interference and utilization in the single bands are linked together as shown by the curves in Figure 6. The collaborative case has an improved reliability, which corresponds to a smaller interference. On the other hand, (P1) optimizes only the thresholds, whereas (P2) requires the contemporary optimization of a nonhomogeneous set of variables (weights and thresholds), yielding inevitably to a more complex problem. Two procedures have been proposed in [8] to solve (P2).

(1) *Sequential optimization* performs at first a spatial optimization that maximizes the *modified deflection coefficient*:

$$d^2_{m,k} = \frac{\mathbf{w}_k^T \boldsymbol{\mu}_1 - \mathbf{w}_k^T \boldsymbol{\mu}_0}{\mathbf{w}_k^T \boldsymbol{\Sigma}_{1,k} \mathbf{w}_k}. \tag{24}$$

Then it performs a *spectral optimization* of the thresholds as if there were only one sensing CR. Sequential optimization performs close to the optimal global solution. The spectral optimization is actually the procedure followed to solve also the thresholds optimization in the noncollaborative case.

(2) *Joint optimization* finds directly thresholds and weights for a global maximization of the throughput. It is optimal in a global sense. Exploitation of hidden convexity needs heavy approximations, so that the final performance is compromised [8].

Then single-band weights optimization (P3) is also analyzed by discussing the throughput graphics.

5.1. Analysis of Nonconvex Classes of Multiband Detection Systems. The multiband frameworks (P1) and (P2) suffer from the same limitations of the convexity constraints. The latter additionally has to be lower bounded when not solved with sequential optimization, which reduces the achievable throughput [8]. The characteristics are analogous for the two problems, because the utilization and interference bounds have the same implications. So the presented results are valid for both frameworks. With the genetic solution the I-R characteristics are calculated for the three CR system classes with the aggregate interference as independent variable. Figure 5 shows four case studies for the problem (P1). The case studies are also presented schematically in Figure 3.

The α_k and β_k values are bounds for the minimum and maximum per-band interference and utilization. By choosing a minimum subband utilization we also impose a minimum interference, and vice versa, for these two quantities are determined by the threshold γ_k. Figure 6 shows the utilization-interference characteristics. If the convexity imposes to operate inside the shaded region then it is impossible to reduce the interference under a certain value, resulting in compromised performance. The region with less utilization but also less unnecessary interference has been excluded before and it is included in the genetic optimization. Convexity exploitation through the utilization constraint $\beta_k \geq 0.5$ is counterproductive for the performance of the system. Therefore comes the importance of a nonconvex maximization tool such as GA.

Different CR systems show different achievable throughput because less or more interference and utilization is permitted in the single bands as shown in Figure 5.

Conservative systems break the convexity with values of utilization β_k below 50%. By reducing β_k we allow transmitting with small bitrates over the subchannels with poor SNR, which are the cause of a high interference. A higher percentage of false alarms is implicitly provoked, but the trade-off is more favorable so that the operative point has a higher throughput and lower interference as β_k decreases. The interested region is mainly for low aggregate interference, whereas, asinthotically, the systems have the same characteristic. The case with $\alpha_k = 0.1$ and $\beta_k = 0.5$ is common in literature for demonstration purposes [8], but it is largely outperformed. *Aggressive* systems with β_k higher than 0.5 show even less favorable characteristics, with low rates at high interferences.

Hostile systems have a maximum interference α_k beyond 50%. By increasing this value we permit more per-band interference with a gain in the throughput. The missed detections increase, so the bitrate increases with a higher slope, because more disturb on the individual bands is permitted. The other systems with low α_k increase with a poorer slope. This interests mainly the higher interferences, whereas small interferences are largely beyond the bound and the operative point remains similar.

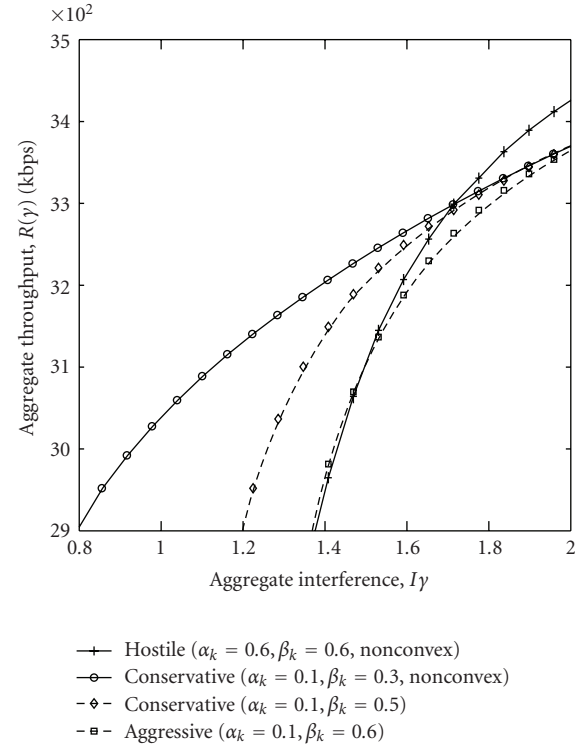


Figure 5: Aggregate opportunistic throughput versus aggregate interference in multiband detection. Convex systems are plotted in dashed line.

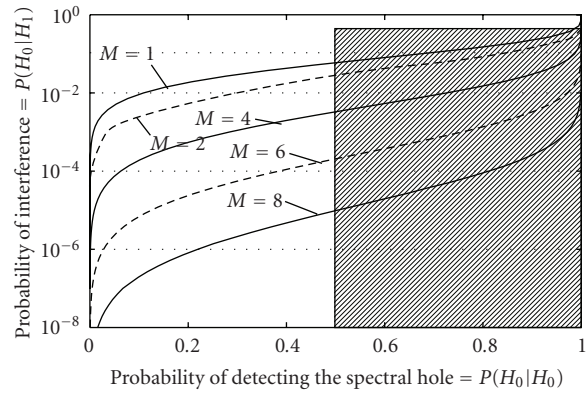


Figure 6: Narrowband transmission against interference probability characteristics in LSC, with $M = [1, 2, 4, 6, 8]$ collaborating CR. The shaded region highlights the convexity domain of (P1).

Aggressive systems may not transmit with low bitrate and low interference because a minimum is imposed to both of them. They pay the mild per-band interference and the acceptable per-band utilization with a worse I-R characteristic.

Figure 7 shows the bar diagrams of the thresholds and the per-band average interference and utilization, for one case of each system in the multiband problem. We notice that the limits for convexity are exceeded. On one hand, more interference is caused in some bands, but with a gain in the throughput that is reflected in the I-R characteristics in Figure 5. On the other hand, less utilization is permitted in other bands, especially those with poor SNR (2nd and 5th bands), but with lower interference. Such systems bring an improvement with respect to the *aggressive* case for higher and lower interferences, respectively.

Not all combinations of α_k and β_k are admitted. Some combinations are unfeasible if the utilization limit β_k implies an interference that does not respect the condition on α_k. Low SNR may support the appearance of such cases just by bringing the Gaussian pdfs of the sensed levels one close to each other.

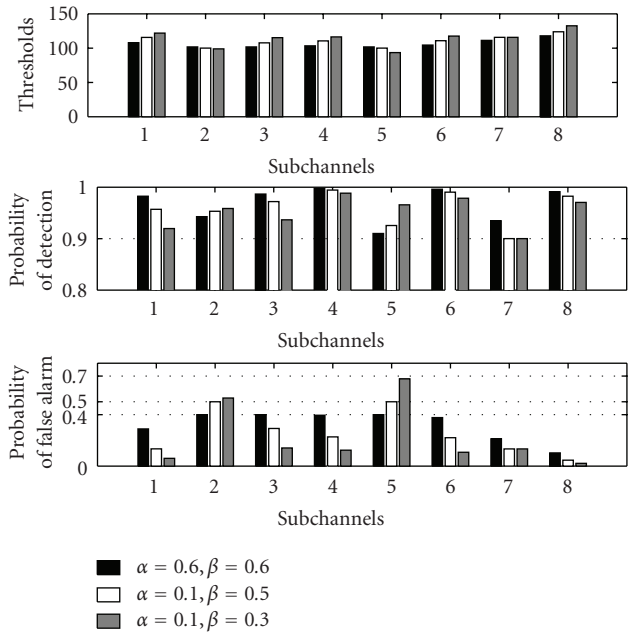


FIGURE 7: Optimized thresholds and probabilities of detection (P_d) and false alarm (P_f) on individual subbands for two of the curves in Figure 5, $\epsilon = 1.3$. SNR = (−3, −5.2, −3.5, −1.9, −6, −2.2, −4, −1.5).

5.2. Analysis of Multiband Detection without Aggregate Constraint. By removing the constraint on the aggregate interference, the problem of maximizing the utilization with a per-band interference constraint has become a series of independent LSC optimizations. Optimum weights are calculated to provide the maximum probability of transmission in each subband with fixed probability of interference. The thresholds are implicit. LSC provides the graphics of transmission-versus-interference probability in Figure 6. The case with one CR is equivalent to a multiband aggregation problem, but the solution is immediate by means of (17). Then the aggregate throughput is calculated afterwards as a linear combination of the subchannel rates. The total rates against the interference to each subchannel system is plotted in Figure 8, which is a direct result of the reliable detection whose characteristics are shown in Figure 6. Besides, LSC is far more simple to be solved, both for GA, that converge easier, and for convex maximization, when working with interference below 50%.

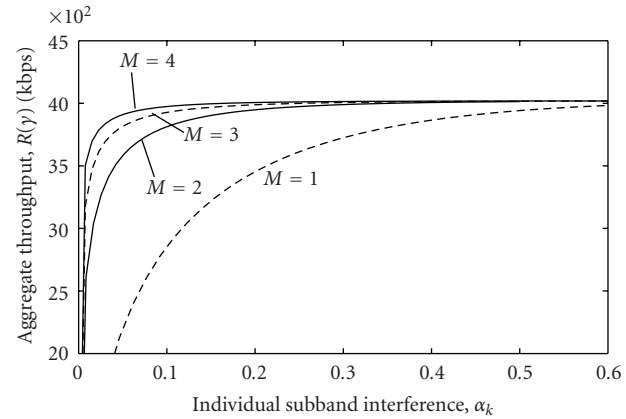


FIGURE 8: Aggregate throughput vs single band interference, with $M = [1, 2, 3, 4]$ collaborating CR.

5.3. Genetic Design. Setting up the parameters of our GA is important to optimize the computation, from the point of view of the expenditure of resources and the convergence precision.

When using a GA, as well as any iterative solving method, a finite difference between the true maximum and the one computed by the GA is expected, because of the finite number of iterations and individual evaluations. The optimization is considered solved when it approaches a negligible error. An error of tens of kilobits (out of some megabits) on the aggregate throughput is considered acceptable. For the aggregate throughput distance in (P1) and (P2) we use a relative measure, the Mean Absolute Percentage Error (MAPE):

$$\text{MAPE}(R) = \frac{1}{Z}\sum_{z=1}^{Z}\left|\frac{\hat{R} - R^{(z)}}{\hat{R}}\right|. \tag{25}$$

From the collaborative side, a squared error of around 10^{-6} on the subband fractional utilization (P_f) is also acceptable. The subchannel utilization error is measured with the Mean Squared Error (MSE):

$$\text{MSE}\left(P_f\right) = \frac{1}{Z}\sum_{z=1}^{Z}\left(\hat{P}_f - P_f^{(z)}\right)^2. \tag{26}$$

$R^{(z)}$ and $P_f^{(z)}$ are the value calculated by the GA during the zth experiment and Z is the number of experiments. $\hat{R}$ and

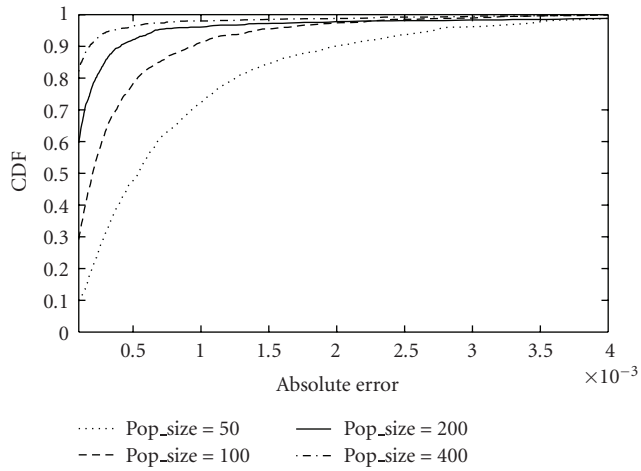


Figure 9: CDF (Cumulative Distribution Function) of the absolute percentage error in thresholds optimization with various population sizes.

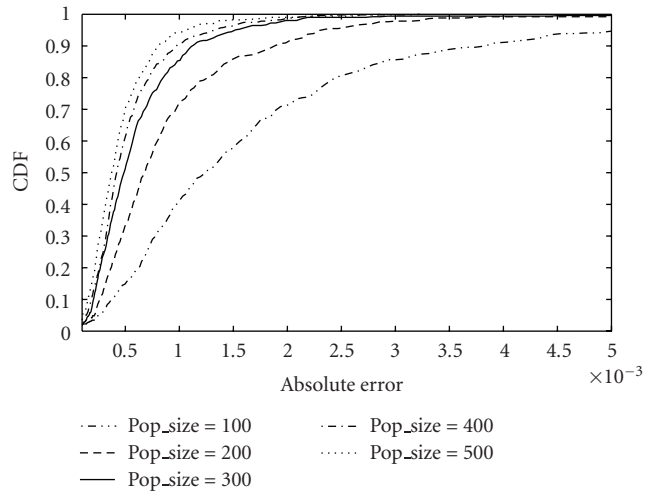


Figure 10: CDF (Cumulative Distribution Function) of the absolute percentage error in joint weights and thresholds optimization with initial points (27) and (28).

$\hat{P}_f$ represent a solution of the maximization several orders of magnitude more precise than the other optimizations. They are obtained at the expenses of a practically unfeasible algorithm but enough accurate to evaluate the precision of the other computations.

We vary the dimension of ordinary elaborations in order to find a compromise between complexity and accuracy. By setting the value of the population size (Pop_size) we control the dimension of the GA in order to evaluate a wider range of *genotypes* and generate a fitter population. By eventually setting the limit on the maximum number of generations (Max _gen) we avoid the algorithm running for a long time before converging.

The mean number of function evaluations to complete an elaboration (N-eval) is our index of complexity which depends directly on the population size and on the number of generations.

We also examine different crossover functions, for a well-chosen crossover criterion converges faster and with higher accuracy.

5.3.1. Spectral Optimization. It consists in the optimization of the thresholds for the noncollaborative problem (P1), as well as for the sequential optimization of (P2) with the weights part solved otherwise (modified deflection coefficient or GA applied to LSC). The characteristics for noncollaborative detection are shown in Figure 5. A MAPE less than 0.1% is enough to infer that the algorithm has converged with acceptable precision. Figure 9 shows the convergence precision while varying the population size in terms of Cumulative Distribution Function (CDF) of the relative error.

5.3.2. Joint Optimization. It consists in the joint optimization of the spatial and spectral variables in (P2). The optimized characteristic is analogous to the one shown in Figure 5. The genes are not homogeneous (weights and thresholds)

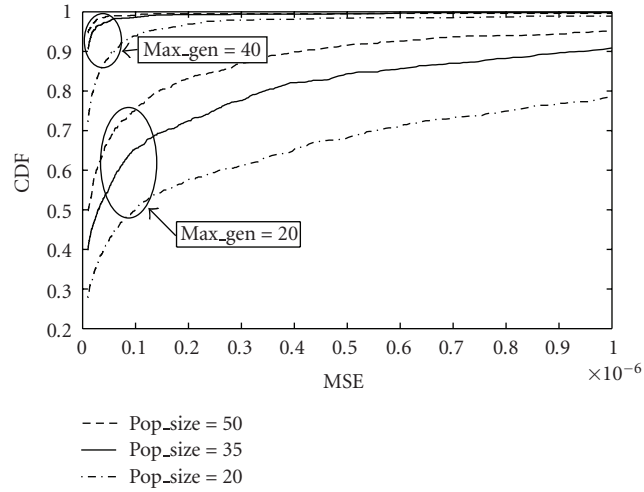


Figure 11: CDF (Cumulative Distribution Functions) of the squared error with various population dimensions in LSC weights optimization.

according to (20), so the evolution results more complex. An expedient for helping the convergence is to set some initial points as the starting population of the GA. We first obtain the initial weights from the maximization of the modified deflection coefficient (with a simple GA or with the procedure in [10]):

$$\mathbf{w}_{k,\text{init}} = \arg\max_{\mathbf{w}} d^2_{m,k}, \quad k = 1,\ldots. \tag{27}$$

Then the initial thresholds are uniformly distributed between the minimum and maximum calculated by means of (11):

$$\boldsymbol{\gamma}_{k,\text{init}} = \Big[\gamma_{\min,k}, \gamma_{\min,k} + \text{step}_k, \gamma_{\min,k} + 2 * \text{step}_k, \ldots, \gamma_{\min,k} + (\text{Pop\_size} - 1) * \text{step}_k, \gamma_{\max,k}\Big]^T, \quad k = 1,\ldots,K, \tag{28}$$

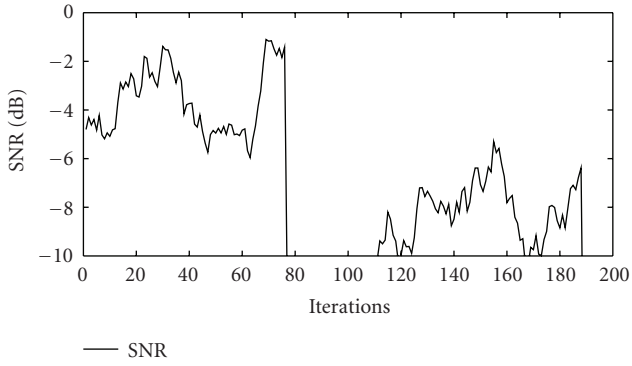


Figure 12: Example of time-variation of the SNR received by each sensor.

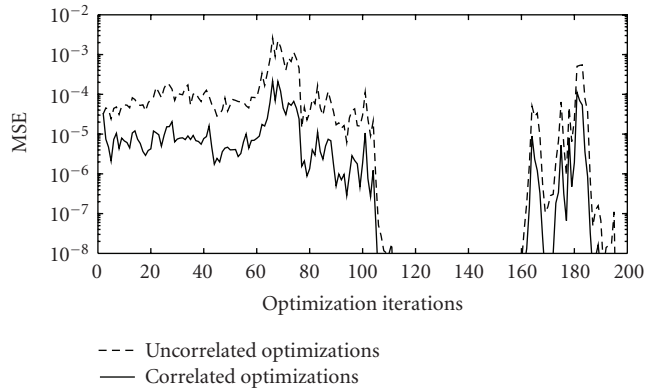


Figure 13: Precision reached with correlated and uncorrelated genetic optimizations of LSC, in terms of Mean Square Error (MSE) with fixed number of generations.

where

$$\text{step}_k = \frac{\gamma_{\max,k} - \gamma_{\min,k}}{\text{Pop\_size}}. \tag{29}$$

With these starting points the GA needs only few generations of joint elaboration to converge to the global optimal solution. Figure 10 shows the CDF of the squared error with different population sizes.

5.3.3. Weights Optimization. It consists in the optimization of the subband weights (18), whose result is shown in Figure 6. It solves the alternative multiband optimization without aggregate constraint (P3). An acceptable convergence is said to be reached with an MSE around 10^{-6}. Figure 11 shows the CDF of the squared error reached by the GA in weight optimization with various dimensions of the population. Since weights optimization is unconstrained, the computational load is exactly $N_{\text{eval}} = \text{Pop\_size} * N_{\text{gen}}$.

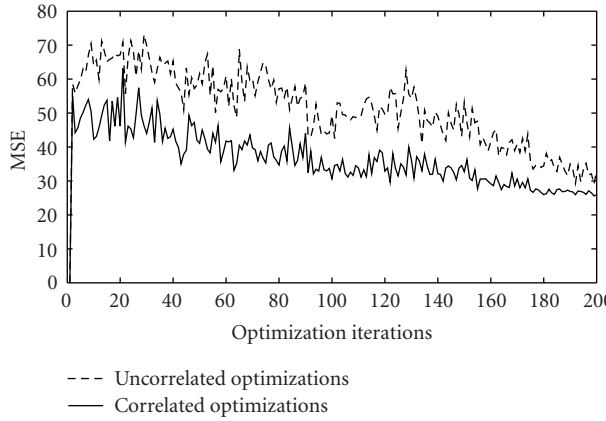


Figure 14: Number of generations necessary to achieve a precision of 10^{-6} with correlated and uncorrelated genetic optimizations of LSC, Pop_size = 20.

6. Variation of the Channel Statistics

Let us consider now the variation of the receiving conditions of the sensors due to the movement. Channel statistics in presence of moving sensors are supposed to vary in time with a certain correlation. A simulated variation in time of the SNR at the radio interface of one CR is shown in Figure 12. The statistic of the received level in presence of transmission $\boldsymbol{\mu}_1$ has lognormal distribution, as it derives from long-term fading and shadowing. The cutoff of the sensor is also simulated, in case of sudden loss of the sensing contribution. A similar variation is followed by each node in the simulation. Measurements correlation is exploited by keeping the result of one elaboration and refining it in the successive instant. Figure 13 shows the precision reached by running the GA for LSC optimization with a fixed number of generations at each instant. The two curves correspond to a genetic optimization from random points every time (memoryless) and from the previous weights as starting vector (with memory). We can see that we gain an order of magnitude of MSE by iteratively updating the objective function (with the channel statistics) and keeping trace of the previous calculated weights. Figure 14 shows instead the number of generations that are needed to reach a certain precision with memoryless optimization using at each instant the previous weights as starting point. In average, 20 generations less are needed to converge, since the weights are tightly correlated between consecutive instants.

7. Conclusion

GAs were proposed as a valid technique for solving the detection problem efficiently and without convexity constraints. The solution is practically analogous to the true mathematical maximum. GAs are able to exploit the correlation of the mobile channel. Unpractical limits due to mathematically unfeasible problems are avoided. *Conservative* systems demonstrate to outperform *aggressive* systems and the throughput increases as we reduce the minimum subband occupancy. In general the complexity of our GA demonstrates to be sustainable and controllable. The computational load does not increase too much as the sensing problem grows or as the GA dimension increases.

References

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40–48, 2008.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[4] "IEEE Standards Coordinating Committee 41 (Dynamic Spectrum Access Networks)," http://www.scc41.org/.

[5] "IEEE 802.22 Working Group on Wireless Regional Area Networks," http://www.ieee802.org/22/.

[6] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 328–337, November 2005.

[7] Z. Tian and G. B. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *Proceedings of the 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '06)*, June 2006.

[8] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, 2009.

[9] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proceedings of IEEE International Conference on Communications (ICC '06)*, pp. 1658–1663, July 2006.

[10] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, 2008.

[11] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Spatial-spectral joint detection for wideband spectrum sensing in cognitive radio networks," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 2793–2796, April 2008.

[12] L. Atzori and A. Raccis, "Network capacity assignment for multicast services using genetic algorithms," *IEEE Communications Letters*, vol. 8, no. 6, pp. 403–405, 2004.

[13] M. Murroni, "Performance analysis of modulation with unequal power allocations over fading channels: a genetic algorithm approach," in *Proceedings of the 14th European Wireless Conference (EW '08)*, Prague, Czech Republic, June 2008.

[14] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, no. 2, pp. 65–85, 1994.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall, Upper Saddle River, NJ, USA, 1998.

[16] Z. Quan, W.-K. Ma, S. Cui, and A. Sayed, "Optimal linear fusion for distributed spectrum sensing via semidefinite programming," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '09)*, pp. 3629–3632, April 2009.

[17] T. Weise, *Global Optimization Algorithms—Theory and Application*, Thomas Weise, 2007.

[18] G. E. Liepins and M. D. Vose, "Deceptiveness and genetic algorithm dynamics," in *Proceedings of the 1st Workshop on Foundations of Genetic Algorithms*, pp. 36–50, July 1990.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 730509, 10 pages
doi:10.1155/2010/730509

Research Article

Mixed-Signal Parallel Compressive Spectrum Sensing for Cognitive Radios

Zhuizhuan Yu,[1] Xi Chen,[1] Sebastian Hoyos,[1] Brian M. Sadler,[2] Jingxuan Gong,[1] and Chengliang Qian[1]

[1] *Department of Electrical and Computer Engineering, The Analog and Mixed Signal Center (AMSC), Texas A&M University, College Station, TX 77843-3128, USA*
[2] *Army Research Laboratory, AMSRD-ARL-CI-CN, Adelphi, MD 20783, USA*

Correspondence should be addressed to Sebastian Hoyos, hoyos@ece.tamu.edu

Received 2 December 2009; Accepted 30 January 2010

Academic Editor: Fred Daneshgaran

Wideband spectrum sensing for cognitive radios requires very demanding analog-to-digital conversion (ADC) speed and dynamic range. In this paper, a mixed-signal parallel compressive sensing architecture is developed to realize wideband spectrum sensing for cognitive radios at sub-Nqyuist rates by exploiting the sparsity in current frequency usage. Overlapping windowed integrators are used for analog basis expansion, that provides flexible filter nulls for clock leakage spur rejection. A low-speed experimental system, built with off-the-shelf components, is presented. The impact of circuit nonidealities is considered in detail, providing insight for a future integrated circuit implementation.

1. Introduction

Cognitive Radio (CR), first proposed in [1], provides a new paradigm to improve spectrum efficiency by enabling Dynamic Spectrum Access (DSA). In CR, *spectrum holes* that are unoccupied by primary users can be assigned to appropriate secondary users as long as the interference introduced by secondary users is not harmful to the primary users [2–4]. The design of cognitive radio networks is a complicated cross-layer procedure [5]. In this paper, we focus on the *spectrum sensing* problem in CR, in which sensing and detection of primary users is done in order to realize Dynamic Spectrum Access.

Spectrum sensing can be a very challenging task for CR due to many factors. First, for the sake of improving the frequency usage efficiency, the sensing bandwidth for CR can expand from hundreds of MHz to several GHz. Second, the sensing radio should be able to detect very weak primary users, which arise due to fading and the hidden terminal problem [5]. With traditional time-domain Nyquist sampling, sensors are needed with both wide bandwidth and high dynamic range, stressing technology, and demanding higher power [6, 7]. Conventional wideband sensing with a high-speed and high-resolution ADC becomes less appealing as the bandwidth becomes significant. Alternative approaches, such as a fixed bank of analog filters followed by parallel ADCs, impose strict requirements on the filter design.

It has been observed that today's spectrum usage presents some *sparsity* in the sense that only a small portion of the available frequency bands are heavily loaded while others are partially or rarely occupied [5]. This frequency usage sparsity can be exploited under the framework of Compressed Sensing (CS) [8, 9] to effectively reduce the sampling rate. The sparse signal can be captured via projection over a random basis that is incoherent with respect to the signal basis, and perfect signal reconstruction from these projections can be obtained with high probability, where the number of random projections is on the order of the signal's information rate rather than the Nyquist rate.

The idea of applying CS for wideband spectrum sensing was reported, for example, in [10]. However, this approach assumes full-rate analog-to-digital conversion which does not reduce the complexity of the spectrum sensing receiver. We have proposed a mixed-signal parallel segmented compressive sensing (PSCS) architecture for wideband spectrum sensing [11], where the high-speed ADCs were avoided by

carrying out an analog basis expansion in parallel before sampling. In this paper we elaborate on the idea of applying the PSCS front-end [11], with special emphasis on implementation issues such as spurious frequency tones, timing, and other mismatches. First, we show that the proposed overlapping windowed integration in the PSCS architecture provides a scheme to mitigate the spurs due to clock leakage by setting the lowpass filter nulls flexibly, which is favorable for practical implementation. Second, a low-speed prototype built with off-the-shelf components is presented in detail from the overall system configuration to building blocks, in which practical constraining issues are addressed.

The remainder of the paper is organized as follows. A brief background on CS is provided in Section 2 and the spectral occupancy signal modeling is given in Section 3. Section 4 introduces the mixed-signal parallel compressive spectrum sensing scheme. Section 5 discusses the spurious frequency rejection schemes in the PSCS front-end. A low-speed prototype is introduced in Section 6. Conclusions are made in Section 7.

2. Compressive Sensing Background

A signal $r(t)$ that is spanned by S basis functions $\Psi_s(t)$ ($s = 1, 2, \ldots, S$), that is, $r(t) = \sum_{s=1}^{S} a_s \Psi_s(t)$, or in the matrix form $\mathbf{r} = \mathbf{\Psi a}$, is a *K-sparse* signal if only K out of the S coefficients $a_s|_{s=1}^{S}$ are nonzero at any time, where $K \ll S$. A signal $r(t)$ is *compressible* if its approximation error by a *K-sparse* signal decays exponentially as K increases.

According to CS theory, a signal that is *sparse* or *compressible* over a known basis $\mathbf{\Psi}$ can be sampled and reconstructed at sub-Nyquist rate, and the sampling rate reduction depends on the signal's sparsity and the reconstruction algorithms. Specifically, the sub-Nyquist rate sampling is achieved by projecting the signal into a transform-domain over which the sampling operation occurs, which is different from the traditional way of sampling the signal in the time-domain. Mathematically, this procedure can be described as $\mathbf{y} = \mathbf{\Phi\Psi a}$, where $\mathbf{y}$ are the collected samples, and $\mathbf{\Phi}$ is incoherent with $\mathbf{\Psi}$ which is the basis for the transform-domain. The reconstruction of the original signal relies on the estimation of the coefficients $\mathbf{a}$, which is obtained by solving the following l_1*–norm* optimization problems, for which many convex optimization techniques or iterative greedy algorithms can be used:

(i) noiseless case:

$$\hat{\mathbf{a}} = \arg\min \|\mathbf{a}\|_1 \quad \text{s.t. } \mathbf{y} = \mathbf{\Phi\Psi a}, \tag{1}$$

(ii) noisy case:

$$\hat{\mathbf{a}} = \arg\min \|\mathbf{a}\|_1 \quad \text{s.t. } \left\|\mathbf{y} - \mathbf{\Phi\Psi a}\right\|_2 \le \epsilon, \tag{2}$$

where ϵ is the error due to the noise.

Note that, in this paper, we generally do not differentiate between *sparse* and *compressible* unless specifically noted.

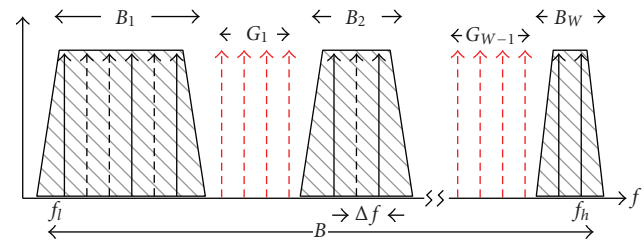


Figure 1: Illustration of the multiband analog signal to the sensing radio.

3. Signal Modeling

The received signal $r(t)$ is a modeled as a multiband analog signal whose spectrum is illustrated in Figure 1. Specifically, we assume that $r(t)$, with a frequency span from f_l to f_h, is the superposition of primary users, perhaps using W different wireless standards [5]. Each wireless standard occupies a certain finite frequency band which consists of multiple channels. According to the measurements done by FCC in the US [12], in many cases the current frequency usage exhibits sparsity because only a part of the allocated channels is utilized at a given time.

Without loss of generality, we assume that $r(t)$ is bandlimited to $[0, f_h]$; so $r(t)$ can be written as

$$r(t) = \int_{-\infty}^{\infty} R(f) e^{j2\pi ft} df = \int_{0}^{f_h} R(f) e^{j2\pi ft} df, \tag{3}$$

where $R(f)$ is the Fourier transform of $r(t)$.

The continuous-time analog signal can be captured with a finite dimensional model; for example, see [13, 14]. We directly approximate $r(t)$ with a model of finite dimension as follows:

$$r(t) \approx \sum_{s=0}^{S-1} R(s\Delta f) e^{j2\pi s\Delta f t} \Delta f, \quad t \in [0, T_s], \tag{4}$$

where $\Delta f = 1/T_s$ is the resolution on the frequency axis and $(S-1)\Delta f = f_h$. In other words, $r(t)$ is approximated as a multicarrier signal bandlimited to $[0, f_h]$ and with a carrier spacing of Δf. The sparse frequency occupancy means that statistically, speaking, only K out of the S carriers are active at any time, where $K \ll S$. The multicarrier model is convenient for representing user occupancy with spectral sparsity. Comparing (3) and (4), we notice that this model is based on a finite dimensional approximation of the signal spectrum. Since there are S unknowns where $R(s\Delta f)$ in (4) and $R(s\Delta f)$ change every T_s seconds, the model in (4) is a case of a Finite Rate of Innovation (FRI) model in which the innovation locations lie on the Nyquist grid. For clarity, we rewrite (4) as

$$r(t) = \sum_{s=0}^{S-1} a_s \Psi_s(t) + n(t), \tag{5}$$

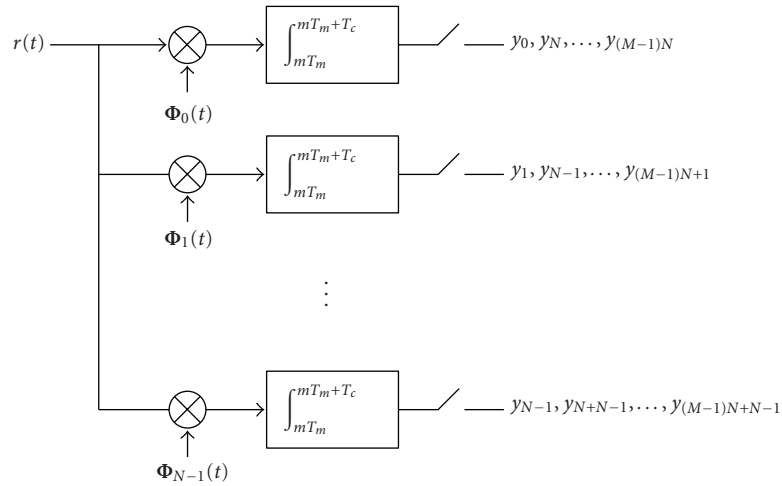


Figure 2: Block diagram of the parallel segmented compressive sensing (PSCS) architecture.

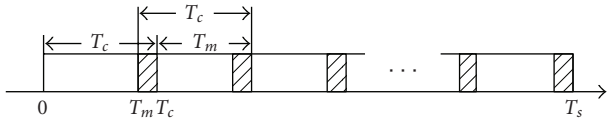


Figure 3: Illustration of overlapping windows.

where $n(t)$ is additive white Gaussian noise (AWGN), $\mathbf{\Psi} = [\Psi_0(t), \Psi_1(t), \ldots, \Psi_{S-1}(t)]$, $\Psi_s(t) = e^{j2\pi s\Delta f t}$, $\mathbf{a} = [a_0(t), a_1(t), \ldots, a_{S-1}(t)] \in \mathbb{C}^S$, $a_s = \Delta f R(s\Delta f)$, and $\mathbf{a}$ has only $K \ll S$ nonzero elements. Since Δf is a scalar, for simplicity, we discard it in the rest of the paper. The spectrum hole detection, for example, energy or feature detection, is usually based on the observed signal spectrum $R(s\Delta f)$, or equivalently, the estimation of the coefficients a_s.

4. Wideband Parallel Compressive Spectrum Sensing

Wideband spectrum sensing is composed of several crucial steps: first, spectrum estimation; second, calculate the sufficient statistics, during which digital signal processing is needed to improve the front-end sensing sensitivity by processing gain and identification of the primary users based on knowledge of the signal characteristics [5]; last, to decide whether there exist primary users based on the sufficient statistics. Here we focus on the wideband spectrum estimation step, that is, estimating the unknown coefficients $\mathbf{a}$ in (5).

4.1. Mixed-Signal Compressive Sensing Architecture. The *parallel segmented compressive sensing (PSCS)* structure is shown in Figure 2, which we first proposed in [11]. For the completeness of this paper, in this section we recap how the analog compressive sensing at sub-Nyquist rate is realized via the PSCS architecture.

In the PSCS architecture, the input signal $r(t)$ is sent to N parallel paths. In the n_{th} path, $r(t)$ is mixed with a random basis function $\Phi_n(t)$. A good choice for the random basis is to use PN (Pseudonoise) sequences because they can be conveniently generated by digital logic circuits. The output of the mixer is then sent to a sliding window with a width of T_c and integrated. Two adjacent windows have an overlapping time $T_c - T_m$, which defines an overlapping percentage OVR $= (T_c - T_m)/T_c$, as shown in Figure 3. The output of the integrators is sampled and M samples are collected at each path. The m_{th} sample of the n_{th} branch is given by

$$y_{mN+n} = \int_{mT_m}^{mT_m+T_c} r(t)\Phi_n^*(t)dt. \tag{6}$$

There are a total of $L = MN$ samples collected every T_s seconds and these samples are organized into a vector as follows:

$$\mathbf{y} = \left[\widetilde{\mathbf{y}}_0^T, \widetilde{\mathbf{y}}_1^T, \ldots, \widetilde{\mathbf{y}}_{M-1}^T\right]^T, \tag{7}$$

where $\widetilde{\mathbf{y}}_\mathbf{m} = [y_{mN}, y_{mN+1}, \ldots, y_{mN+N-1}]^T$ is the vector consisting of the m_{th} samples from all N branches.

Similarly, we can calculate the *reconstruction matrix* $\mathbf{V} = \mathbf{\Phi\Psi} = \{v_{i,j}\}_{L\times S}$. The element at the $mN + n$ row and the s column is given by

$$V_{mN+n,s} = \int_{mT_m}^{mT_m+T_c} e^{j2\pi s\Delta f t}\Phi_n^*(t)dt. \tag{8}$$

Therefore, we have $\mathbf{y} = \mathbf{Va}$. Then, we can estimate $\mathbf{a}$ by solving the problem in (1) and reconstruct the original signal using $\mathbf{r} = \mathbf{\Psi}\hat{\mathbf{a}}$.

4.2. A Wideband Spectrum Sensing Example. To show the effectiveness of the proposed wideband PSCS architecture we present a simulation, where the input signal is modeled as a frequency-domain sparse multi-carrier signal as given in (5). The mixed-signal compressive sensing based on the PSCS architecture given in Figure 2 is used for spectrum

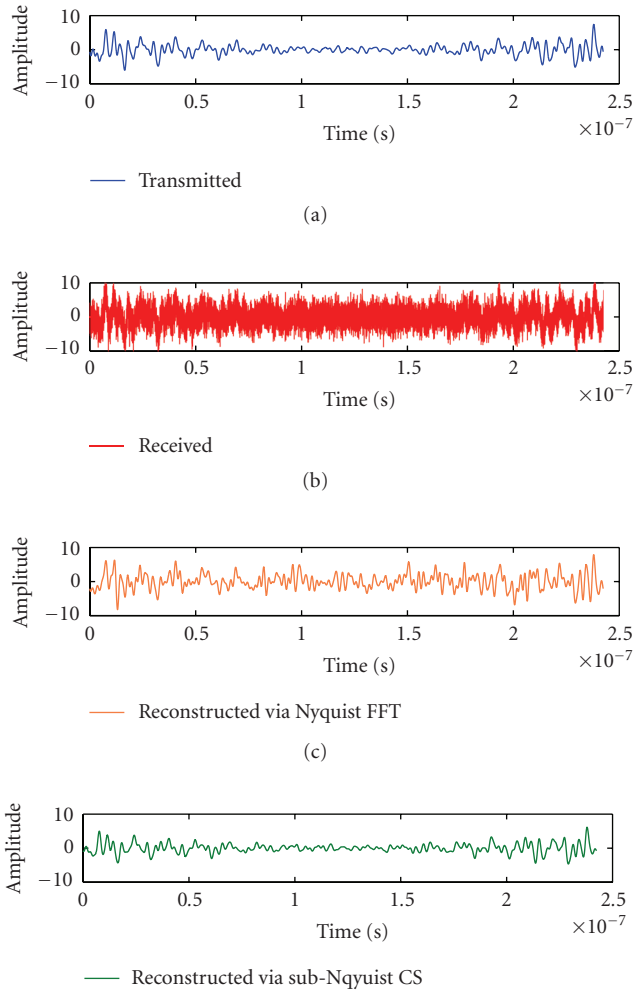


FIGURE 4: Time-domain signals of a simulated multiband signal. From top to bottom, the four plots represent the transmitted signal by primary users, the received primary users' signal at the sensing radio, the reconstructed signal from the time-domain samples via the Nyquist rate ADC, and the reconstructed signal from the transform-domain samples via mixed-CS at an NSR of 0.32.

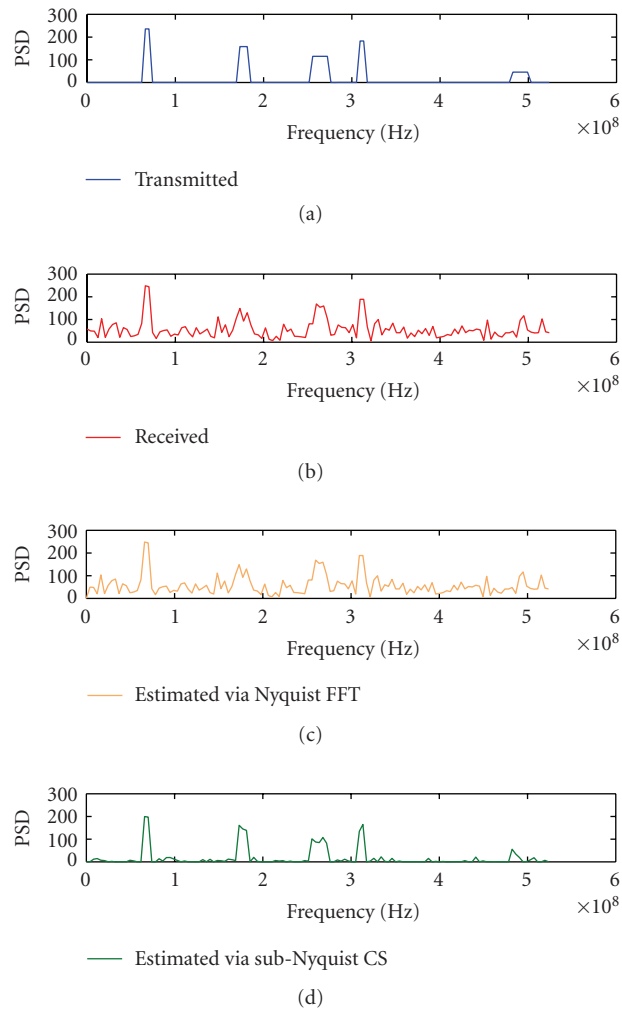


FIGURE 5: Frequency-domain signals of a simulated multiband signal. From top to bottom, the four plots represent the transmitted signal by primary users, the received primary users' signal at the sensing radio, the reconstructed signal from the time-domain samples via the Nyquist rate ADC, and the reconstructed signal from the transform-domain samples via mixed-CS at a NSR of 0.32.

estimation. The sampling rate reduction is measured by the Normalized Sampling Rate (NSR), which is defined as

$$\text{NSR} = \frac{f_{CS}}{f_{Nq}} = \frac{MN}{S}, \tag{9}$$

where f_{CS} is the sampling rate required using the PSCS and f_{Nq} is the corresponding Nyquist sampling rate. The signal reconstruction quality is evaluated by the normalized Mean Square Error (MSE), which is equal to

$$\frac{\|\mathbf{a} - \hat{\mathbf{a}}\|_2}{\|\mathbf{a}\|_2}. \tag{10}$$

In the simulation, the input signal to the PSCS architecture is assumed to be a *17-sparse* frequency-domain multi-carrier signal with 128 subcarriers, that is, $S = 128$ and $K = 17$. There are 5 primary bands with an overall bandwidth of 528 MHz. The subcarrier spacing $\Delta f = 528\,\text{MHz}/128 = 4.125\,\text{MHz}$ and the primary user's frequencies are $(17, 18, 43, 44, 45, 63, 64, 65, 66, 67, 76, 77, 118, 119, 120, 121, 122) \times 4.125\,\text{MHz}$. The input power dynamic range of the primary users is 15 dB. $\text{SNR}_{\text{overall}} = 0\,\text{dB}$, where $\text{SNR}_{\text{overall}}$ is the total signal power over the whole bandwidth divided by the total noise power over the whole bandwidth. (Note how noisy the received signal is in this example, shown in Figure 4.) In Figures 4 and 5, from top to bottom, the four plots represent the primary transmitted signal, the received primary users' signal at the sensing radio, the reconstructed signal from the time-domain samples via the Nyquist rate ADC, and the reconstructed signal from the transform-domain samples via mixed-CS at an NSR of 0.32. The measured MSE for the two reconstructed signals is -5 dB and -14 dB, respectively. Note that even with a lower sampling rate, the sensing radio

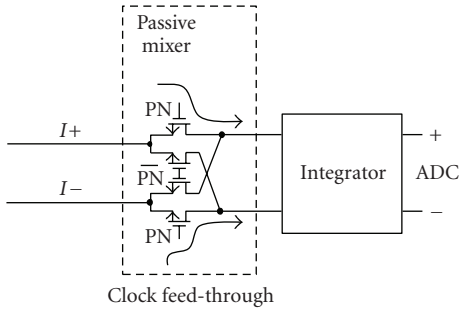


Figure 6: Clock leakage into the integrators from the clock of the PN generators.

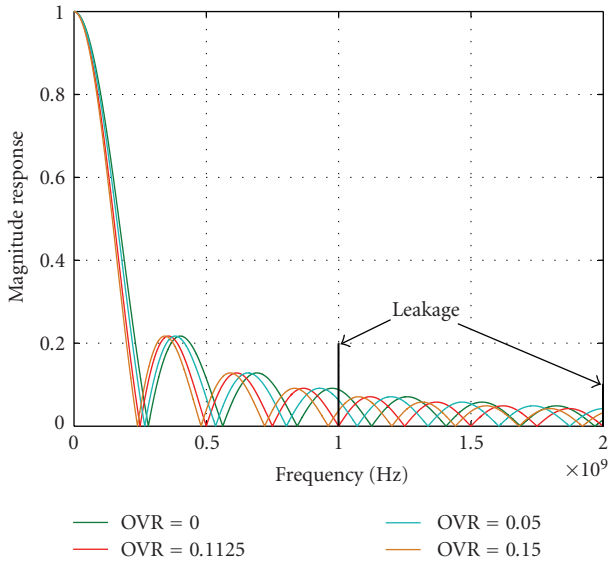


Figure 7: The location of the spurious leakage frequency relative to the filter nulls with different overlapping ratio. With OVR = 0, the strongest clock leakage is close to the peak of the filter's 3rd sidelobe; with OVR = 0.1125, the strongest clock leakage is on the 4th null of the filter.

based on mixed-signal PSCS is more robust against noise than the traditional digital approach based on the DFT, because CS takes advantage of the knowledge of the signal structure and its sparsity.

5. Flexible Spur Rejection via the Overlapping Windowed Integration

In addition to the capability of sensing and reconstructing sparse signals at sub-Nqyuist rate, the PSCS architecture has many special characteristics. For example, the parallel architecture gives a design tradeoff between the sampling rate and the system complexity [11]. In this section, we focus on the PSCS architecture's spurious frequency rejection schemes. Since one critical type of spur in the PSCS architecture is the leakage of the clocks for the PN generators to the integrator, as illustrated in Figure 6, we will focus on this particular type of spur in this section, although the rejection scheme applies more generally.

Recall that in Figure 2, the output after the mixer is sent to a sliding window with a width of T_c and integrated over T_c seconds, and there is an overlap time of $T_c \times \mathrm{OVR}$ between two adjacent windows as illustrated in Figure 3. The integrator, with a reset every T_c seconds, provides a simple realization of a sinc type lowpass filter with nulls at frequencies of $f_0 \times k$, where $f_0 = 1/T_c$. By setting the random generator clock frequency equal to a harmonic of the reset frequency, the sinc nulls coincide with spur frequencies from the random generator clock and so filters them, where the overlapping scheme provides the flexibility on setting the locations of the nulls. In some cases, without the overlapping scheme, the objective of setting the clock frequency on the nulls of the sinc type lowpass filter may conflict with the sampling rate requirement which is determined by the signal's sparsity. In order to show this, consider the following example.

Let the input signal to the PSCS architecture be a *19-sparse* frequency-domain multi-carrier signal with 128 subcarriers, that is, $S = 128$ and $K = 19$, which corresponds to a sparsity of 15%. The subcarrier spacing is $\Delta f = 1\,\mathrm{GHz}/128 = 7.8125\,\mathrm{MHz}$ and the symbol duration time is $T_s = 1/\Delta f = 128$ nanoseconds. The locations of the K active subcarriers are chosen randomly and changed every T_s seconds. According to simulation results, the minimum NSR is $0.5625 = 72/128$ for this parameter setup. Equivalently speaking, 72 samples are needed per 128ns to reconstruct the signal perfectly. Using two parallel paths, 36 samples are collected every 128 nanoseconds at each path, that is, $M = 36$ and $N = 2$. With this parameter setup and without the overlapping scheme, $T_c = T/M = 128/36 = 3.56\,\mathrm{ns}$, $f_0 = 1/T_c = 281.25\,\mathrm{MHz}$, and the nulls of the sinc type lowpass filter occur at $k \times 281.25\,\mathrm{MHz}$.

There may exist some leakage into the integrators from the clock signal, as illustrated in Figure 6. According to the CS theory, the clock frequency is usually at the Nyquist frequency f_{Nq} where $f_{Nq} = 1\,\mathrm{GHz}$ in this example. Because $f_{Nq}/f_0 \approx 3.56$, the spurs due to the clock leakages will fall near the 3rd sidelobe's peak of the sinc type lowpass filter and bring distortion to the reconstructed signal. With the overlapping scheme, we can choose $T_c = 4\,\mathrm{ns}$ and $f_0 = 250\,\mathrm{MHz}$ by introducing an overlapping ratio of 11.43%, then $f_{Nq}/f_0 = 4$ and the spurs due to the clock leakage can be filtered (considering the clock resolution requirement, an overlapping ratio of 11.25% is suggested in practice.). Based on Figure 3, this can be mathematically expressed as

$$T = T_c(M - (M-1)\mathrm{OVR}), \tag{11}$$

$$f_{Nq} = \frac{S}{T}, \tag{12}$$

$$f_0 = \frac{1}{T_c}, \tag{13}$$

$$\Longrightarrow \frac{f_{Nq}}{f_0} = \frac{S}{M - (M-1)\mathrm{OVR}}. \tag{14}$$

According to (14), given a desired sampling rate, or equivalently speaking, a specific M, varying OVR will change the

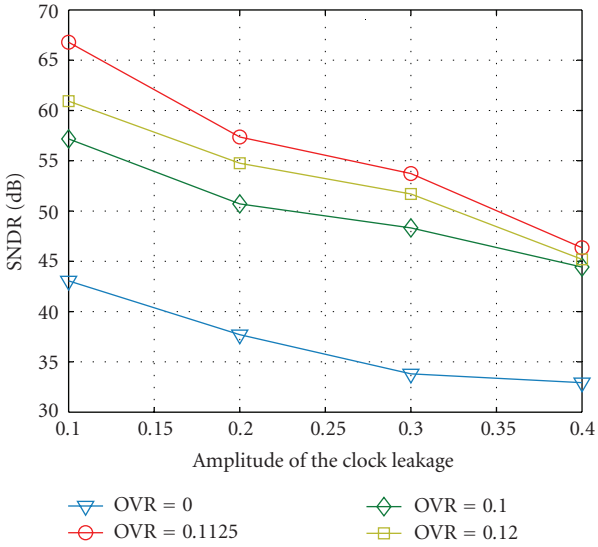


Figure 8: MSE of the reconstructed signal with different overlapping ratio when there is clock leakage, where the input signal is an 18-sparse multi-carrier signal with 128 possible subcarriers and sampled by the 2-path PSCS working at 56.25% of the Nyquist rate.

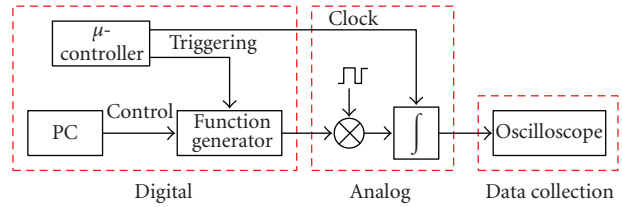


Figure 9: Overall configuration of the prototype using off-the-shelf components.

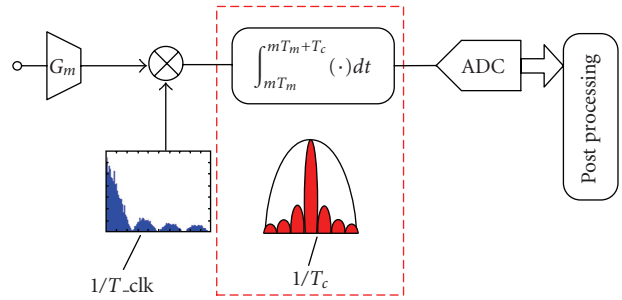


Figure 10: Macro model of one parallel path.

relative location of the leakage frequency to the filter nulls, as illustrated in Figure 7.

Note that if we do not want to introduce any overlapping but still wish to null out the clock leakage, the only option in the above example is to increase the sampling rate and make f_{Nq}/f_0 an integer no less than $\lfloor 3.56 \rfloor$. By introducing a nonzero OVR, we can conveniently make f_{Nq}/f_0 an integer without increasing the sampling rate.

Figure 8 shows the MSE of the reconstructed signal versus the overlapping ratio when there is some clock leakage into the integrators. Note that in the simulation the amplitude of each subcarrier is set to 1. Since $K = 19$, the signal's peak amplitude is 19. Allowing a 10 dB margin to account for the multi-carrier signal's large peak-to-average ratio, the clock leakage with an amplitude of 0.1 (0.4) is roughly 35 dB (23 dB) below the signal's average power. As shown in Figure 8, the flexibility of setting the null frequencies by the overlapping scheme can bring about 20 dB gain after filtering the spurs due to the clock leakage.

Note also that the overlap in the integration windows provides wider filter nulls than the sinc filter. Because of the existence of the phase noise on the clock signal in practice, even if we can set the clock on the null frequency, it is inevitable that remains some leakage due to the widening of the spurs spectrum. The wider nulls provides the possibility of further improving the harmonic rejection when the phase noise is significant.

6. Low-Speed Off-the-Shelf Component Prototype

As a proof of concept, we built a low-speed prototype using off-the-shelf components, where the input signal is a real BPSK modulated multi-carrier signal with 4 active subcarriers and the active subcarriers hop over the frequencies $(i * 2 - 1)$ KHz $(i = 1, 2, \ldots, 100)$ every 500 microseconds. Considering the system complexity, we employed 4 parallel paths for the prototype. Simulation shows that the signal can be reconstructed perfectly when each parallel path produces 16 samples every 500 microseconds, which corresponds to 32% of the Nyquist sampling rate.

6.1. Overall Configuration. The overall configuration of the prototype is shown in Figure 9, where the digital part is responsible for generating the input sparse signal, the triggering signal, the pseudorandom basis, and the clock. The analog part is used to realize the random basis projection that is essential for the signal reconstruction. The built-in ADC in the oscilloscope is used to collect the sampled data. Then, the collected data is sent to a PC and processed via Matlab code to reconstruct the signal. In the following sections, each building block will be introduced in detail.

6.2. Multicarrier Signal Generator. An Agilent 33120A arbitrary waveform generator is used to generate the input multitone sparse signal. Specifically, the multitone signal is programmed in the PC first and then downloaded into the wave generator. The output port of the generator is triggered by the microcontroller in order to synchronize with the integrator clock that is also generated by the microcontroller.

6.3. Mixers and Integrators. Figure 10 depicts the macro-model of one path in the prototype. As shown, the input signal is first translated into current by the OTA and then mixed with the pseudorandom signal. After mixing, the signal is integrated in the sampling capacitor with a timing window. In the sampling circuit the interleaving capacitor is employed. Finally the ADC yields digital output data. The OTA we employed is a TIOPA861 with Gm of 116mS

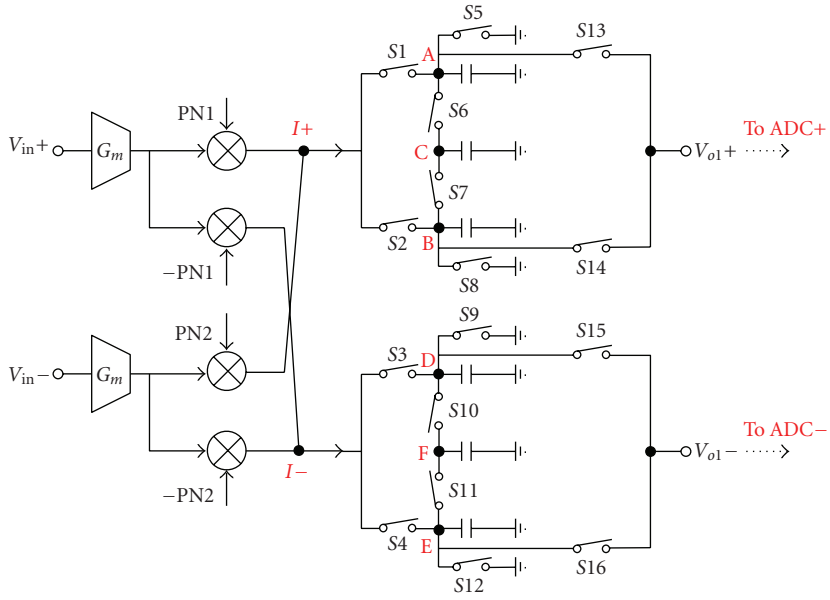


Figure 11: Circuit implementation of one parallel path.

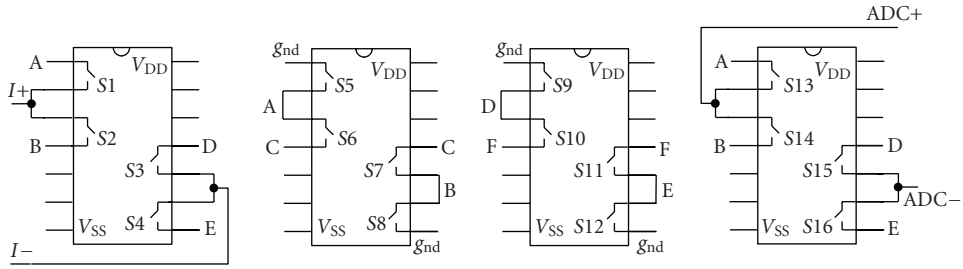


Figure 12: Pin connection of the integrator in one parallel path.

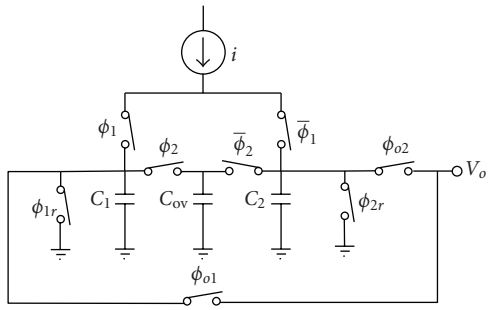


Figure 13: Schematic of the integrator with overlapping.

and all the switches are implemented with transmission gate CD4066BCN.

The pseudorandom number (PN) is -1 or 1, whose spectrum is a sinc function. The main lobe is from 0 to $1/T_{\text{clk}}$, where T_{clk} is the clock period of the PN generator. In our test bed T_{clk} is 1μs. After the mixing, the signal is shaped by the embedded lowpass filter provided by the integration window. The frequency response of the LPF is a sinc function. The main lobe spans $1/T_c$, where T_c is the integration time. In our test bed the $1/T_c$ is roughly 30 KHz.

The random projection of the input analog signals is realized with mixers and integrators. Figure 11 gives the circuit implementation of one parallel path and Figure 12 gives the corresponding pin connection relationship for the integrator. The transconductance amplifier (G_m stage) translates the signal voltage into current, which can be easily mixed with the pseudorandom numbers ($1/-1$) by the following passive switch mixer. After mixing, the signal is integrated with an overlapping window and then sampled by the ADC in each path. The circuit is built up differentially so that the system is more robust to supply noise, clock jitter, and even-order harmonics. The double balanced passive mixer does not introduce significant noise and distortions.

At each path, the mixer consists of transmission-gate switches controlled by PN sequences. The PN sequence is implemented with a linear feedback shift register (LFSR). In our prototype, the clock frequency is chosen to be 1 MHz, which is higher than the Nyquist sampling rate. Because the PN sequences are repeated every 500 μs and there are 4 parallel paths, we need 4 independent PN sequences with a length of 500. An 11-bit LFSR is used to generate a PN sequence with a length of 2047 and then divided into 4 segments. As a check, the autocorrelation function of the PN sequences is calculated to make sure that the four PN sequences are incoherent.

An overlapped time-interleaving charge-domain sampling integrator is chosen for the analog path. The

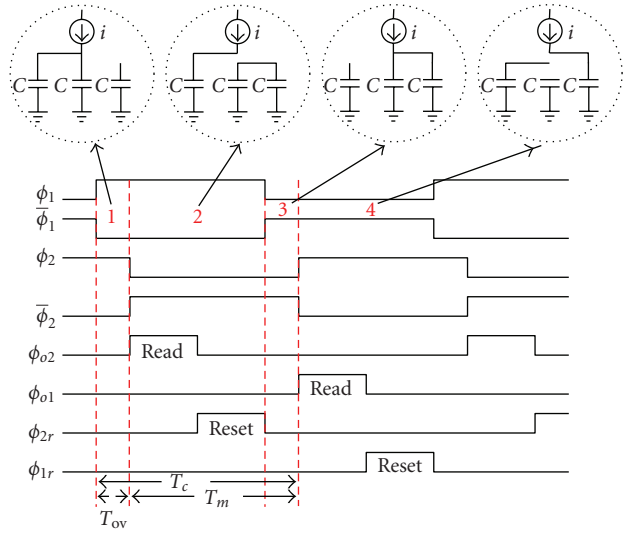


Figure 14: Operation of the interleaved overlapping windowed integration and the related clocks.

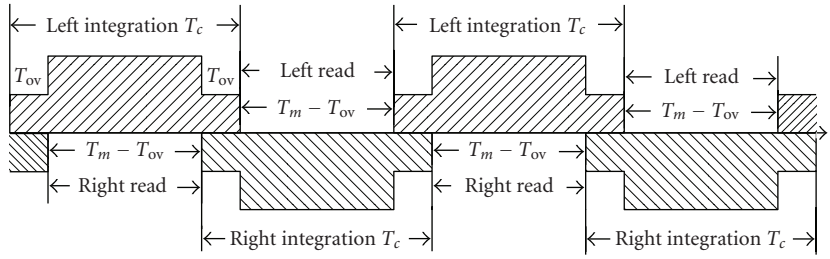


Figure 15: Illustration of the overlapping windowing in the prototype.

integrator schematic is shown in Figure 13, ϕ_1 and $\overline{\phi}_1$ are two integration switches for the left and right branches, respectively. ϕ_{o1} and ϕ_{o2} are readout switches; ϕ_{r1} and ϕ_{r2} are reset switches. By utilizing these six switches combined with the two integration capacitors C_1 and C_2, according to the clock diagram shown in Figure 14, we can realize a conventional time-interleaving charge-domain integrator without overlapping. Time interleaving means when the left branch is integrating while the right is reading out, and vice versa. By doing this, a complete sampling of the signal is achieved. In addition to time interleaving, a small overlapping time is introduced by one more capacitor C_{ov} and two control switches ϕ_2 and $\overline{\phi}_2$.

As shown in Figure 14, phase1 and phase3 are to realize the overlapping through charge redistribution and sharing, and phase2 and phase4 are the readout times for the right and the left branches, respectively. During phase1, the input current charges both C_1 and C_{ov} while C_2 is idle. Since all capacitors have the same value, the current splits equally by half into both capacitors. In the succeeding phase, C_{ov} is switch-connected to C_2 and readout together, so that C_{ov} is integrating for the right branch during phase1. Equivalently, as shown in the timing window diagram, the window splits by half during the overlapping time. The key point here is that both branches are integrating and no data is readout during window overlapping times.

Note that the overlapping windowing realized using the circuit in Figure 13 is somewhat different from the overlapping windowing in Figure 3, as shown in Figure 15. In Figure 3, the charges accumulated during the current window period include 100% of the charges from the last T_{ov} seconds of the previous windowing period but no charges from the next windowing period. In Figure 13, the charges accumulated during the current windowing period include 50% of the charges from both the last T_{ov} seconds of the previous windowing period and the first T_{ov} seconds of the next windowing period, which is more realistic from the implementation perspective.

6.4. Data Collection and Signal Reconstruction. For simplicity, we use the inherent ADC of the oscilloscope (Tectronix TDS 3054 500 MHz, 5 Gs/s) to sample the output of the integrators. The sampled data is transferred to the PC via the GBIP port. With the collected samples, the signal is reconstructed as described in Section 2.

6.5. Dealing with Circuit Nonidealities. While implementing the prototype, it is inevitable that the system has some nonidealities such as the delay caused by each component, the gain variation, and the mismatch among parallel paths. Considering all the nonideal factors, the actual relationship

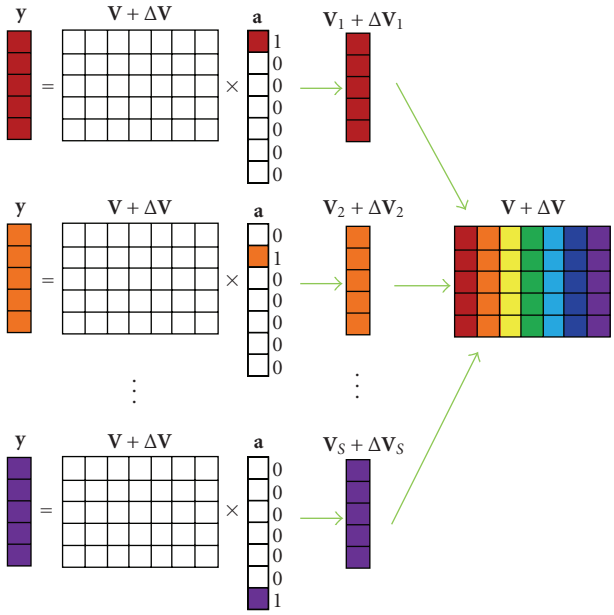


Figure 16: Illustration of the direct training approach to deal with the circuit imperfections.

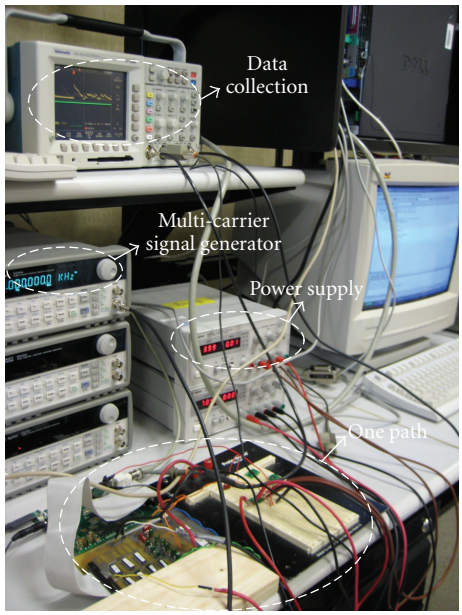


Figure 17: Testing setup of the prototype.

between the collected samples $\mathbf{y}$ and the coefficients $\mathbf{a}$ becomes

$$\mathbf{y} = \tilde{\mathbf{V}}\mathbf{a} = (\mathbf{V} + \delta\mathbf{V})\mathbf{a}, \tag{15}$$

where the element at the $mN + n$ row and the s column of $\tilde{\mathbf{V}}$ is given by

$$\tilde{V}_{mN+n,s} = \int_{mT_m+\delta t_1}^{mT_m+T_c+\delta t_2} \alpha e^{j2\pi(s\Delta f+\delta f)t+\theta}(\Phi_n(t) + \delta\Phi_n(t))dt. \tag{16}$$

Table 1: Testing results of the prototype.

Subcarrier's amplitude (mV)	Subcarrier frequencies of the input testing signal (kHz)	Subcarrier frequencies of the reconstructed signal (kHz)
0.3	[+61, +121]	[+61, +121]
0.3	[+41, +131]	[+41, +131]
0.3	[+61, −131]	[+61, −131]
0.3	[−51, +63, +111]	[−51, +63, +111]
0.2	[+71, −85, +91, −101]	[+71, −85, +91, −101]

Here, δt_1 and δt_2 reflects the timing error on the slicing window, δ_f reflects the frequency offset, α and θ reflects the gain and phase mismatches, and the $\delta\Phi_n(t)$ reflects the error of the random basis which could be attributed to the jitter and nonzero response time.

Because the actual relationship between $\mathbf{y}$ and $\mathbf{a}$ is given by (15), we need to replace $\mathbf{V}$ with $\tilde{\mathbf{V}}$ in (1) when estimating $\mathbf{a}$; otherwise, some extra error will be introduced. In [15], the authors discussed the impact of some circuit imperfections, such as the finite settling time of the PN sequences, and the timing uncertainty, and a background calibration algorithm based on LMS was proposed to compensate for the error due to these circuit nonideal factors. Because of the complexity of the background calibration, here we use a more simple approach based on direct training to deal with the circuit nonidealities. The direct training approach is illustrated in Figure 16. During the training stage, we inject a single-tone signal one at a time to the prototype and collect the samples from the 4 parallel paths, so that these samples will fill one column of the reconstruction matrix $\tilde{\mathbf{V}}$. After sending 100 single-tone signals, we obtain a complete matrix which will be used for signal reconstruction.

This pilot-based method is based on the assumption that the system is linear and time-invariant. Fortunately, our circuit level design ensures that the input signal swing is within the linear range of the system, and the microcontroller ensures that the system has the same initial condition for every run. Therefore, the linear time-variant assumption is reasonable. Implementing the background calibration for circuit imperfection compensation is part of our future work.

6.6. Testing Results. The testing setup for the prototype is shown in Figure 17. A series of experiments are done to test the functionality of the system. Table 1 summarizes the testing results, where + and − stand for the polarity of the BPSK modulation. Note that we scale the amplitude of each subcarrier according to the number of tones such that the amplitude of the multi-carrier signal is within the dynamic range of the system. From the testing results, the prototype achieves the design specification.

7. Conclusions

The Parallel Segmented Compressive Sensing (PSCS) front-end is able to sample and reconstruct analog *sparse* and *compressive* signals at sub-Nqyuist rate. The overlapping

windowed integration in the PSCS front-end provides a spurious frequency rejection scheme by setting the lowpass filter nulls on the spurious frequencies without sacrificing the sampling rate requirement. A low-speed prototype is built with off-the-shelf components, which is able to sense sparse analog signals at sub-Nyquist rate.

Acknowledgments

The authors acknowledge the contributions of the students, faculty, and sponsors of the Analog and Mixed-Signal Center at the Texas A&M University. This research project was funded under the DARPA Analog-to-Information Receiver Development Program (Army Research Laboratory Cooperative Agreement no. W911NF-08-2-0047) and the Army Research Laboratory Cooperative Technology Alliance (ARL-CTA) contract no. DAAD-19-01-2-0011. Some of the results in this paper were reported in the conference of IEEE ICASSP'08, Las Vegas, Neveda, USA [11] and IEEE DCAS'09, Dallas, Texas, USA [15].

References

[1] J. Mitola III, *Cognitive radio: an integrated agent architecture for software defined radio*, Ph.D. dissertation, Royal Institute of Technology (KTH), May 2000.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[4] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, 2007.

[5] D. B. Cabric, *Cognitive radios: system design perspective*, Ph.D. dissertation, University of California, Berkeley, Calif, USA, December 2007.

[6] E. A. M. Klumperink, R. Shrestha, E. Mensink, V. J. Arkesteijn, and B. Nauta, "Polyphase multipath radio circuits for dynamic spectrum access," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 104–112, 2007.

[7] R. H. Walden, "Performance trends for analog-to-digital converters," *IEEE Communications Magazine*, vol. 37, no. 2, pp. 96–101, 1999.

[8] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[10] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '07)*, vol. 4, pp. 1357–1360, Honolulu, Hawaii, USA, April 2007.

[11] Z. Yu, S. Hoyos, and B. M. Sadler, "Mixed-signal parallel compressed sensing and reception for cognitive radio," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '08)*, pp. 3861–3864, Las Vegas, Nev, USA, March 2008.

[12] FCC, "Spectrum policy task force report," Tech. Rep. 02-135, ET Docket, 2002.

[13] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: compressed sensing for analog signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 993–1009, 2009.

[14] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2986–2997, 2009.

[15] Z. Yu and S. Hoyos, "Digitally assisted analog compressive sensing," in *Proceedings of the IEEE Dallas Circuits and Systems Workshop*, 2009.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 985458, 13 pages
doi:10.1155/2010/985458

Research Article

Cross-Layer Throughput Optimization in Cognitive Radio Networks with SINR Constraints

Miao Ma and Danny H. K. Tsang

Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Correspondence should be addressed to Miao Ma, eemma@ust.hk

Received 22 September 2009; Revised 10 February 2010; Accepted 14 April 2010

Academic Editor: Massimiliano Laddomada

Recently, there have been some research works in the design of cross-layer protocols for cognitive radio (CR) networks, where the Protocol Model is used to model the radio interference. In this paper we consider a multihop multi-channel CR network. We use a more realistic *Signal-to-Interference-plus-Noise Ratio (SINR)* model for radio interference and study the following cross-layer throughput optimization problem: (1) Given a set of secondary users with random but fixed location, and a set of traffic flows, what is the max-min achievable throughput? (2) To achieve the optimum, how to choose the set of active links, how to assign the channels to each active link, and how to route the flows? To the end, we present a formal mathematical formulation with the objective of maximizing the minimum end-to-end flow throughput. Since the formulation is in the forms of *mixed integer nonlinear programming (MINLP)*, which is generally a hard problem, we develop a heuristic method by solving a relaxation of the original problem, followed by rounding and simple local optimization. Simulation results show that the heuristic approach performs very well, that is, the solutions obtained by the heuristic are very close to the global optimum obtained via LINGO.

1. Introduction

Cognitive radio technology [1–3] provides a novel way to solve the spectrum underutilization problem. In cognitive radio (CR) networks, there are two types of users: primary users and secondary users. A primary user is the rightful owner of a channel, while a secondary user periodically scans the channels, identifies the currently unused channels, and accesses the channels opportunistically. The secondary users organize among themselves an ad hoc network and communicate with each other using these identified available channels. As a result, a multihop multichannel CR network is formed. How to efficiently share the spectrum holes among the secondary users, therefore, is of interest.

In this paper, we are interested in studying the opportunistic spectrum sharing problem among the secondary users, but our concern is on a cross-layer design of spectrum sharing and routing with SINR constraints. The main issues we are going to address include the following.

(1) Given a set of secondary users with random but fixed location, and a set of traffic flows, what is the max-min achievable throughput?

(2) To achieve the optimum, how to choose the set of active links, how to assign the channels to each active link, and how to route the flows?

There have been some research works on cross-layer protocols in CR networks. Hou et al. [4] characterized the behaviors and constraints for a cognitive radio network from multiple layers, including modeling the spectrum sharing and subband division, scheduling and interference constraints, and flow routing. Shi and Hou [5] developed a formal mathematical model for scheduling feasibility under the influence of power control; the formulation is a cross-layer design optimization problem encompassing power control, scheduling, and flow routing. Subsequently, on the basis of the work in [5], Shi and Hou [6] implemented their cross-layer optimization framework in a distributed manner and compared the performance of the distributed

optimization algorithm with the upper bound and validated the efficacy. The work in [4–6] assume that the links are unidirectional, and to avoid collision only the designated receiver is need to be out of the interference caused by another transmitter. Ma and Tsang [7] proposed a cross-layer design on spectrum sharing and power control, where bidirectional links were considered and all nodes were operated at an optimal common power level at which the total spectrum utilization is maximized. Ma and Tsang [8] also proposed a cross-layer design on spectrum sharing and routing, where the channel heterogeneity (which is a unique feature for cognitive radio) was considered and modeled.

In the previous work, however, a common limitation exists since all such cross-layer protocols [4–8] are designed on the basis of the Protocol Model for radio interference, where the interference range is assumed to be limited and no interference is caused beyond the interference range. As a result, in the Protocol Model the conflict relationships among the wireless links are *binary*. However, in reality the *aggregate* interference of a large number of far transmitters could be significant and may cause interference on a receiver, and a near transmitter may not necessarily cause interference on a receiver if the transmitter properly controls its transmission power. Therefore, a definite criticism of the Protocol Model is that interference is not a binary relationship [9–11].

In order to solve the above realistic problems, the *Signal-to-Interference-plus-Noise Ratio (SINR)* model is adopted. The rationale of SINR model is to compare the SINR with the *additive* interference calculation at the receiver with a threshold. Some researchers have adopted the SINR model when they consider the link scheduling, power control, or throughput improvement and and so forth. in wireless networks. For example, Brar et al. [11] investigated throughput improvements in wireless mesh networks by replacing CSMA/CA with an STDMA scheme where transmissions were scheduled according to the SINR model. Chafekar et al. [12] studied a cross-layer latency minimization problem in wireless networks with SINR model for interference. Behzad and Rubin [13] developed a new mathematical programming formulation for minimizing the schedule length in multihop wireless networks while meeting the requirements on the SINR at intended receivers.

In this paper, we consider a multihop multi-channel CR network. We adopt the (more realistic) *Signal-to-Interference-plus-Noise Ratio (SINR)* model to study the wireless channel interference. Different from the work in [4–6], we consider the links being bidirectional because we believe the link level acknowledgments in an ad hoc network are a must. We propose a cross-layer optimization framework which jointly considers the spectrum sharing and routing with SINR constraints. The optimization problem is in the forms of a *mixed integer nonlinear programming (MINLP)* and the objective is to maximize the minimum end-to-end flow throughput. Since the MINLP formulation is NP-hard in general, we present a heuristic methodology by solving a relaxation of the original problem, followed by rounding and simple local optimization. Simulation results show that the heuristic approach works very well; that is, the solutions obtained by the heuristic are very close to the global optimum obtained via LINGO [14].

The rest of this paper is organized as follows. In Section 2, we describe the assumptions and system model. Section 3 introduces two interference models: one is protocol model and the other is SINR model. Section 4 presents the cross-layer design of spectrum sharing and routing with SINR constraints, and the formulation is in the forms of a mixed integer nonlinear programming (MINLP) problem. The heuristic approach is proposed in Section 5 to solve the MINLP problem. Section 6 presents the simulation results. Finally, Section 7 concludes the paper.

2. Assumptions and System Model

We consider a cognitive radio (CR) network with n secondary users, denoted by the set V and the cardinality $|V| = n$. There are M orthogonal channels in the network, denoted by the set C and the cardinality $|C| = M$. Each secondary user individually detects the available channels, and the set of available channels that can be used for communication is different from node to node. Let C_i denote the set of available channels observed by node i, and we have $C_i \subseteq C$.

Each secondary user i (where $1 \leq i \leq n$) has a programmable number of radio interfaces, denoted by γ_i. We assume that the radio interface is able to tune in a wide range of channels, but at a specific time each radio interface can only operate on one channel [15].

2.1. Static Node Location with a Centralized Server. We assume that the node locations are static. We also assume the set of available channel at each secondary user is static. This corresponds to the applications with a slow varying spectrum environment (e.g., TV broadcast bands). We assume that there exists a centralized server in the CR network. Each secondary user reports its location and the set of available channels to the spectrum server. The spectrum management and flow routing, therefore, is simple and coordinated. Note that the formulations in the work [4, 5, 8] are also centralized and for static scenario (i.e., both node location and set of available channels at each node are static). Table 1 lists the notations used in this paper.

2.2. Bidirectional Links. We consider bidirectional links, rather than unidirectional links, due to two reasons [16].

(1) Wireless medium is lossy. We cannot assume that a packet can be successfully received by a neighbor unless the neighbor acknowledges it. In an ad hoc network, the link level acknowledgments are necessary.

(2) Medium access controls such as IEEE 802.11 implicitly rely on bi-directionality assumptions. For example, a RTS-CTS exchange is usually used to perform virtual carrier sensing.

Thus, if node i can transmit data to node j and vice versa, then we represent this by a (bidirectional) link, denoted by

Table 1: Notations.

Symbol	Meaning
V	Set of secondary users
n	Number of secondary users $\lvert V\rvert$
E	Set of potential links
G	Network graph
C	Set of available channels
M	Number of available channels $\lvert C\rvert$
E_i	Set of links incident on node i
C_i	Set of available channels at node i
C_e	Set of available channels at link e
γ_i	Number of radio interfaces at node i
r_i	Transmission range for node i
R_i	Interference range for node i
Δ	Guard zone
E'	Set of links that have simultaneous transmissions with link e by using a same channel
E'_m	Set of links that have simultaneous transmissions with link e by using channel m
P	Common transmission power
N_i	Noise power at node i
I_i	Interference power at node i
N	Noise power at each node
β	The minimum threshold to decode a transmission at a receiver
d_{ij}	Distance between nodes i and j
G_{ij}	Propagation gain from node i to node j
η	Path loss exponent
x_e^m	Binary variable indicating if link e is active on channel m
B_e^m	Capacity of link e by using channel m
T	Minimum end-to-end flow throughput
W	Bandwidth of a channel
Q	Number of communication sessions
$s(q)$	Source node for session q
$d(q)$	Destination node for session q
$f_{i,j}^q$	Traffic flow from i to j for session q
$f_{j,i}^q$	Traffic flow from j to i for session q

$e = (i, j)$, between node i and node j. Moreover, we let C_e denote the set of available channels for the link e, and we have $C_e = C_i \cap C_j$.

2.3. Common Transmission Power. According to the study by Narayanaswamy et al. [16], to ensure that links are bidirectional, the simplest approach is to assume that nodes are homogeneous; that is, nodes transmit at the same power. In this paper we assume that each secondary user is equipped with an omnidirectional antenna. Similar to [16], we also assume that each secondary user transmits at the same power. Note that this assumption is used in [4] as well.

A *bidirectional* link, denoted by $e = (i, j)$, can be established between nodes i and j if there exists a transmission power P under which the *Signal-to-Noise Ratio (SNR)* in the absence of cochannel interference at nodes i and j is not less than a threshold β, that is,

$$\frac{P \cdot G_{ij}}{N_j} \geq \beta, \qquad \frac{P \cdot G_{ji}}{N_i} \geq \beta, \tag{1}$$

where β is signal-to-noise Ratio (SNR) threshold, G_{ij} (and G_{ji}) denotes the channel propagation gain from i to j (and from j to i), and N_j (and N_i) denotes the noise power at node j (and node i). Since it has been commonly assumed that G_{ij} is equal to G_{ji} [4–8, 17–21], and N_i is equal to N_j, we make the same assumptions here and thus we have

$$\frac{P \cdot G_{ij}}{N} \geq \beta. \tag{2}$$

We let E_i denote the set of links incident on node i, which can be obtained by

$$E_i = \left\{ e : \frac{P \cdot G_{ij}}{N} \geq \beta \right\}. \tag{3}$$

Let E denote the union of E_i; we have

$$E = \bigcup_{i \in V} E_i. \tag{4}$$

As a result, we obtain an *undirected* connectivity graph $G = (V, E)$ to represent the CR network, where V is the set of secondary users denoted by the vertices of the graph, and E is the set of edges between two vertices (i.e., secondary users).

3. The Interference Model

In wireless networks, there are three types of interference: *duplexing interference*, *primary interference,* and *secondary interference*. In this paper, we assume that links using different channels do not interfere with each other. *Interference* only occurs among the links sharing *the same channel*.

The *duplexing interference constraint* [22] only prohibits any node from simultaneously transmitting and receiving on any frequency band (i.e., the case in Figure 1 is not allowed).

The *primary interference constraint* prohibits any node from simultaneously transmitting or receiving on any band (i.e., neither case in Figures 2(a), 2(b) nor 2(c) is allowed). In other words, links that shared a common node cannot transmit or receive simultaneously on any channel. Obviously, the duplexing constraint is less stringent than the primary interference constraint. And also, the duplexing and primary interference constraints are applicable to the links which share a common node (see Figures 1 and 2), and particularly, *these constraints hold irrespective of the interference model*.

The *secondary interference constraint*, on the other hand, further prohibits any node from transmitting when a neighbor node within its interference range is receiving from another node. Different from the duplexing and primary interference constraints, the secondary interference

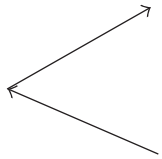


Figure 1: Duplexing interference constraint.

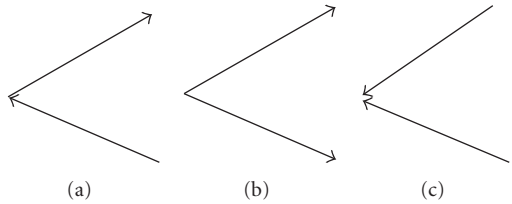


Figure 2: Primary interference constraint.

constraints are applicable to those links which do not share a common node (see Figure 8 shown in the appendix for better understanding).

For the purpose of modeling the secondary interference, there are two models [23, 24]: the Protocol Model and the Physical Model. Since rationale of the Physical Model is based on the SINR at the receiver, we call the Physical Model as the SINR model hereinafter in this paper. The relationship of these three types of interferences is shown in Figure 3.

3.1. The Protocol Model. Gupta and Kumar proposed the Protocol Model [23] which implicitly assumed that links are unidirectional. With this assumption, collisions only occur when the designated receiver is interfered by another transmitter. Basically, the Protocol Model assumes that the interference range is limited and no interference will be caused beyond the interference range. We let r_i and R_i denote the transmission range and interference range for any node i, respectively; then we have $R_i = (1+\Delta)r_i$, where Δ is the guard zone to prevent a neighboring node from transmitting on the same channel at the same time [23]. The Protocol Model claims that a transmission from node i to node j is successful if and only if any node k which may cause interference on node j (i.e., if $d_{kj} \le R_k$ where d_{kj} denotes the distance between k and j) is not simultaneously transmitting.

A more realistic version, however, assumes that IEEE 802.11 MAC is employed and thus the links are bidirectional (due to RTS-CTS and ACK exchange). We usually call this version as 802.11-style Protocol Model [7, 8]. Suppose that link $e = (i, j)$ and link $e' = (k, h)$ are established and both are bidirectional. They are also active on a same channel. The 802.11-style Protocol Model states that a transmission on link e between nodes i and j is successful if and only if for any link $e' = (k, h)$, such that $d_{kj} \le R_k$ or $d_{ki} \le R_k$ or $d_{hj} \le R_h$ or $d_{hi} \le R_h$, is not simultaneously transmitting.

Note that the Protocol Model leads to *binary* conflict relationships among the wireless links. In other words, any two links either interfere with each other or can be active simultaneously, regardless of the other ongoing signal transmissions.

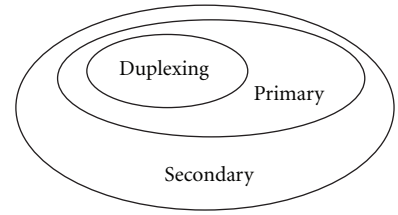


Figure 3: Relationship of three types of interferences.

3.2. The SINR Model. As we mentioned before, in reality the aggregate interference of a large number of far transmitters could be significant and may cause interference at the receiver, and a near transmitter may not necessarily cause interference at the receiver if the transmitter properly controls its transmission power. Thus, the main limitation of the Protocol Model is that interference is not a binary relationship. These problems can be overcome by means of the SINR model, whose rationale is as follows.

3.2.1. Unidirectional Links. Before we consider bidirectional link, let us first consider unidirectional link. For clarification, we let $e = i \to j$ and $e' = k \to h$ denote two unidirectional links and suppose that they are active on a same channel. The transmission from node i is successfully received by node j if and only if the SINR at the receiving node j is not less than a threshold β. That is,

$$\frac{P \cdot G_{ij}}{I_j + N} \ge \beta, \tag{5}$$

where I_j denotes the interference power at node j. To calculate I_j, we need to sum all the links $\{e' = k \to h\}$ that have simultaneous transmissions with link e on a same channel. Therefore we have

$$\frac{P \cdot G_{ij}}{P \cdot \sum_{k \to h \in E'} G_{kj} + N} \ge \beta, \tag{6}$$

where E' contains all links that have transmissions concurrent with link e by using a same channel. The SINR model accurately captures the fact that interference is caused by *aggregate* effect of the simultaneous active links.

3.2.2. Bidirectional Links. Next, we extend the SINR model from unidirectional link to bidirectional link. To distinguish from the unidirectional links, we let $e = (i, j)$ and $e' = (k, h)$ denote two bidirectional links and suppose that they are active on a same channel. Because the interference raised by node k and node h might be different, we need to choose the maximum one. To ensure the transmission on link e between nodes i and j to be successful, the SINR at both nodes i and j is not less than a threshold β. We also need to sum all the links $\{e' = (k, h)\}$ that have simultaneous transmissions with link e on a same channel. To the end, we obtain

$$\frac{P \cdot G_{ij}}{P \cdot \sum_{(k,h) \in E'} \max\left(G_{kj}, G_{hj}, G_{ki}, G_{hi}\right) + N} \ge \beta. \tag{7}$$

Note that the SINR model is more accurate than the Protocol Model since it better captures the physical propagation. Moreover, in the SINR model a correct packet reception is allowed even in the presence of (moderate) interference, and the cumulative character of interference is taken into account. The main drawback of this model lies in its high complexity, as the interference is described as the complex mathematical relationships.

4. Cross-Layer Design of Spectrum Sharing and Routing

In this section, we present a cross-layer optimization framework which jointly designs the spectrum sharing and routing. Spectrum sharing can be done either in time domain or in frequency domain. In this paper, we consider frequency domain channel assignment. Spectrum sharing is to determine which link is going to be active and which channel will be assigned to each active link, and our target is to form a *conflict-free topology*. Routing is to determine which path each traffic flow is going to travel from the source node to the destination node. We allow *multipath* for each traffic flow. Different from the previous work, in this paper, we adopt the *SINR model* for radio interference and consider links being *bidirectional*.

4.1. Link Assignment. We say that link e is active only if there is a transmission on channel m over link e. We define a 0-1 binary variable x_e^m as follows:

$$x_e^m = \begin{cases} 1 & \text{if link } e \text{ is active on channel } m, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

4.2. Interference Constraints. In this paper, we consider both primary and secondary interference constraints. We term the secondary interference constraints as the *SINR constraints* hereinafter in this paper, since we use the SINR model to model them.

In the remaining part of this subsection, we let $e = (i, j)$ denote a link and let $e' = (k, h)$ denote another link. Both links are active and use a same channel m for transmission.

4.2.1. Primary Interference Constraints. By using a same channel, each node can either transmit or receive but not both, at a given time. In other words, links that share a common node cannot transmit or receive simultaneously on any channel.

For ease of presentation and also for notational convenience, *each link is also understood as a set of two nodes*; then we define

$$e \cap e' \begin{cases} = \emptyset & \text{if } i \neq k \text{ and } j \neq k \text{ and } i \neq h \text{ and } j \neq h, \\ \neq \emptyset & \text{if } i = k \text{ or } j = k \text{ or } i = h \text{ or } j = h. \end{cases} \tag{9}$$

Clearly, we use the notation $e \cap e' = \emptyset$ to denote that the two links e and e' do not share a common node, and the notation $e \cap e' \neq \emptyset$ to denote that the two links e and e' share a common node.

Thus, the primary interference constraint can be expressed as follows:

$$\begin{aligned} x_e^m + x_{e'}^m &\leq 1 \\ (m \in C_e \cap C_{e'}, e' \cap e \neq \emptyset, e' \neq e, e' &\in E, e \in E). \end{aligned} \tag{10}$$

4.2.2. SINR (i.e., Secondary Interference) Constraints. For the links that do *not* share a common node but share a common channel, they are applicable to the SINR constraints if the links are active simultaneously. That is, a transmission on a bidirectional link $e = (i, j)$ is successful if and only if the SINR at either node i or node j is not less than the minimum required threshold β. This leads to the following constraint:

$$\begin{aligned} PG_{ij}x_e^m - \beta P \sum_{e' \in E'_m} \max\left(G_{kj}, G_{hj}, G_{ki}, G_{hi}\right) x_{e'}^m - \beta N \\ \geq \Upsilon(x_e^m - 1) \quad (m \in C_e \cap C_{e'}, e \cap e' = \emptyset, e' \in E, e \in E), \end{aligned} \tag{11}$$

where E'_m contains all links that have transmissions concurrent with link e by using channel m. Similar to [13], we introduce a sufficiently large positive number Υ in the constraint (11), where the constraint becomes "redundant" when link e is not active (i.e., $x_e^m = 0$). Notice that we only sum the interference caused by those *active* links because link e' will not cause any interference on link e whenever link e' is not active (i.e., $x_{e'}^m = 0$). And also note that when we calculate the interference caused by the active link $e' = (k, h)$, we choose the maximum interference caused from either node k or node h to either node i or node j (due to bidirectional link).

4.3. Node-Radio Constraint. A node can establish multiple links with its neighboring nodes if it can tune each of its radio interface to a different channel. However, the number of established links at each node is constrained by the number of its radio interfaces. This leads to the following constraint:

$$\sum_{e \in E_i} \sum_{m \in C_e} x_e^m \leq \gamma_i \quad (i \in V). \tag{12}$$

4.4. Multipath Routing Constraints. We consider multiple traffic flows in the network. We term the traffic flow for each source-destination pair as a communication session and use q ($q = 1, 2, \ldots, Q$) to index each session. Let $s(q)$ and $d(q)$ represent the source node and destination node for session q. Because the links are bidirectional, the traffic flow on each link can be in either direction. Thus, for any link $e = (i, j)$, we let $f_{i,j}^q$ (and $f_{j,i}^q$) denote the traffic flow traveling from i to j (and from j to i) for the session q, where $(i, j) \in E, i \neq j$. For each traffic flow, we allow *multi-path* routing.

Our definition of the maximum throughput is *max-min flow rate* [25]. That is, our target is to maximize the minimum end-to-end flow throughput that can be achieved

in the network. Therefore, the multi-path routing constraints are listed as follows:

$$f_{i,j}^{q} \geq 0, \quad f_{j,i}^{q} \geq 0 \quad ((i,j) \in E, q \in Q), \tag{13}$$

$$\sum_{\substack{k:k \neq d(q),\\ (k,i) \in E_i}} f_{k,i}^{q} = \sum_{\substack{j:j \neq s(q),\\ (i,j) \in E_i}} f_{i,j}^{q} \tag{14}$$

$$(i \neq d(q), i \neq s(q), i \in V, q \in Q),$$

$$\sum_{j:(i,j) \in E_i} f_{i,j}^{q} \geq T \quad (i = s(q), q \in Q), \tag{15}$$

$$\sum_{j:(j,i) \in E_i} f_{j,i}^{q} \geq T \quad (i = d(q), q \in Q), \tag{16}$$

$$\sum_{q \in Q} f_{i,j}^{q} + \sum_{q \in Q} f_{j,i}^{q} \leq \sum_{m \in C_e} B_e^m \quad (e = (i,j), e \in E), \tag{17}$$

where T is the minimum end-to-end throughput for every session.

The constraint (13) restricts the amount of flow on each link to be nonnegative. The constraint (14) states that at each node, except the source node and destination node, the amount of incoming flow is equal to the amount of outgoing flow. The constraint (15) represents that the minimum outgoing flow from each source node is at least T. The constraint (16) states that the minimum incoming flow to the destination node is at least T. The constraint (17) indicates that the sum of the flows over all sessions traversing a link cannot exceed the link capacity.

To calculate the link capacity, we let W denote the bandwidth of each channel, and let B_e^m denote the capacity of link e by using channel m. Assuming Gaussian noise and interference, we have

$$B_e^m = W \log_2 \left(1 + \frac{PG_{ji} x_e^m}{P \cdot \sum_{e' \in E'_m} \max\left(G_{kj}, G_{hj}, G_{ki}, G_{hi}\right) x_{e'}^m + N} \right) \quad (m \in C_e \cap C_{e'}, e \cap e' = \varnothing, e' \in E, e \in E). \tag{18}$$

4.5. Problem Formulation. We aim to maximize the minimum end-to-end throughput, and this optimization problem can be formulated as

$$\max \quad T \tag{19}$$

$$\text{Subject to:} \quad x_e^m = 0, 1 \quad (m \in C_e, e \in E), \tag{20}$$

$$x_e^m + x_{e'}^m \leq 1 \quad (m \in C_e \cap C_{e'}, e' \cap e \neq \varnothing, e' \neq e, e' \in E, e \in E), \tag{21}$$

$$PG_{ij} x_e^m - \beta P \sum_{e' \in E'_m} \max\left(G_{kj}, G_{hj}, G_{ki}, G_{hi}\right) x_{e'}^{\mathrm{m}} - \beta N \geq \Upsilon(x_e^m - 1) \quad (m \in C_e \cap C_{e'}, e \cap e' = \varnothing, e' \in E, e \in E), \tag{22}$$

$$\sum_{e \in E_i} \sum_{m \in C_e} x_e^m \leq \gamma_i \quad (i \in V), \tag{23}$$

$$f_{i,j}^{q} \geq 0, \quad f_{j,i}^{q} \geq 0 \quad ((i,j) \in E, q \in Q), \tag{24}$$

$$\sum_{\substack{k:k \neq d(q),\\ (k,i) \in E_i}} f_{k,i}^{q} = \sum_{\substack{j:j \neq s(q),\\ (i,j) \in E_i}} f_{i,j}^{q} \tag{25}$$

$$(i \neq d(q), i \neq s(q), i \in V, q \in Q),$$

$$\sum_{j:(i,j) \in E_i} f_{i,j}^{q} \geq T \quad (i = s(q), q \in Q), \tag{26}$$

$$\sum_{j:(j,i) \in E_i} f_{j,i}^{q} \geq T \quad (i = d(q), q \in Q), \tag{27}$$

$$\sum_{q \in Q} f_{i,j}^{q} + \sum_{q \in Q} f_{j,i}^{q} \leq \sum_{m \in C_e} B_e^m \quad (e = (i,j), e \in E), \tag{28}$$

where γ_i is constant and B_e^m can be obtained by (18). x_e^m (binary integer) and $f_{i,j}^{q}$ and $f_{j,i}^{q}$ (rational number) are decision variables. The objective function is a linear function; however, (28) is a nonlinear constraint. The optimization problem is in the form of *mixed integer nonlinear programming* (MINLP) problem and can be solved by LINGO.

5. Finding the Optimal Solution by Heuristic

We are interested in finding the optimal solution under which the minimum end-to-end flow throughput is maximized. However, the original problem is in the form of MINLP since the constraint (28) is nonlinear due to the logarithm function. But if we investigate the formulation more carefully, we find that the complexity of the MINLP problem formulation does *not* lie in the nonlinear logarithm function in constraint (28). Instead, the complexity comes from the existence of the binary variable x_e^m. The reason is that as long as the values of x_e^m are determined (i.e., the set of active links and the channel assignment on each active link are determined), then this MINLP reduces to an LP, which can be solved in polynomial time. To this end, we develop a heuristic method by solving a relaxation of the original problem, followed by rounding and simple local optimization [26].

5.1. Relaxation. We start by relaxing the MINLP problem to the following format.

$$\max \quad T \tag{29}$$

$$\text{Subject to:} \quad 0 \leq x_e^m \leq 1 \quad (m \in C_e, e \in E) \quad \text{constraints} \quad (21)\text{–}(28). \tag{30}$$

That is, we allow the variables x_e^m to take values between 0 and 1. The relaxed problem can be solved in polynomial time. By solving the relaxed problem, we obtain an upper bound of the optimal value of the original problem, and we let $\mathbf{X}^*$ denote the relaxed solution that produces the upper bound.

5.2. Getting Independent Sets. In order to determine the set of active links and form a conflict-free topology, we need to obtain the independent sets (i.e., the set of links that can be simultaneously active on a channel). The reason is that we can significantly speed up the search process by combining the independent sets together with the rounding and local optimization (introduced below).

To obtain the independent sets, Karnik et al. [25] proposed a smart enumerative technique. In this paper, we extend this technique into a more general case from the following two aspects. (1) In [25], all nodes are assumed to transmit at a *single* channel. But in our formulation we consider a more realistic scenario where each node (i.e., secondary user) is able to access a *set* of available channels, and especially, the set of available channels is *different from node to node*. (2) In [25] the links are *unidirectional*, but in our formulation the *bidirectional* links are considered.

For this technique, similar to [25] we make the following additional assumptions. (Interested readers are advised to refer [25] for details on why the above three assumptions are reasonable.)

(A1) The propagation gains are modeled by isotropic path loss. That is, the propagation gain from node i to node j is

$$G_{ij} = \left(\frac{d_0}{d_{ij}}\right)^{\eta}, \tag{31}$$

where d_0 is the far-field crossover distance and η denotes the path loss exponent.

(A2) The minimum distance for any pair of nodes is $d_{\min}$.

(A3) The nodes are located in a square size $L \times L$ area.

Theorem 1. *Under the assumptions (A1)–(A3), the number of simultaneous transmissions on a same channel (i.e., the size of maximum independent sets on a channel) is upper bounded by* $2L^2/\pi(d_{\min}\beta^{1/\eta})^2$.

Proof. Please see the appendix. □

Section 6 will show the extent of complexity reduction by using this technique together with rounding and local optimization.

5.3. Rounding. The next step is to round the relaxed problem solution $\mathbf{X}^*$ to a valid binary integer solution $\hat{\mathbf{X}}$. To create $\hat{\mathbf{X}}$, we can simply round the one (say x_e^m) with the largest value to 1. According to the independent set, with $x_e^m = 1$ we can immediately decide some variables which share the same channel m with the link e to be 1 or 0. After fixing some decision variables to 1 or 0 in the first iteration, we update a new relaxed LP for the second iteration. We can solve this new LP, then again round the one with the largest value to 1, and set some additional variables to 0 accordingly. The iteration continues and eventually we can determine all $\{x_e^m\}$ to either 0 or 1.

Upon fixing all the x_e^m values, the original MINLP reduces to an LP problem, which can be solved in polynomial time. It is worth emphasizing that, unlike the solutions obtained by relaxation, the final solution obtained here is a *feasible* solution since all x_e^m values are binary instead of rational numbers.

5.4. Local Optimization. Further improvement can be obtained by a local optimization method, starting from $\hat{\mathbf{X}}$. Suppose that for channel m there are n_m independent sets, and we use $1, 2, \ldots, n_m$ to index each independent set. Since, in the initial solution $\hat{\mathbf{X}}$, one of n_m independent sets is active on channel m, then we use v ($1 \le v \le n_m$) to index the active independent set. Then for channel m, we observe its independent sets and cycle through $k = 1, 2, \ldots, (n_m - 1)$ while $k \ne v$, and at the kth step replacing the kth independent set as 1. If this change leads to an improvement for the objective function, we accept the change and continue. Otherwise we go on to the next independent set of channel m. We continue until we have tried all the independent sets for channel m. The same process repeats for all the channels. Numerical experiments show that this local optimization method can lead to significant improvement on the objective function.

6. Simulation Results

In this section, we present simulation results for our heuristic method and compare it with the upper bound and the global optimum. The upper bound is obtained by solving a relaxation of the original problem, while the global optimum is obtained by LINGO which is a mathematical software package. The default settings for the simulations are as follows. The noise power at every receiver is equal to -100 dBm. η and d_0 are taken to be 4 and 0.1 m, respectively. The minimum threshold (β) is set to 2.3 dB.

We consider two scenarios: one is regular topology and the other is random topology. We make no claims that these topologies are representative of typical cognitive radio networks. The reason that we have chosen these two simple topologies is to facilitate detailed discussion of the results and for the illustration purpose. For the propagation model, we adopt the isotropic path loss shown in (31). However, we stress that the validity of the conclusions drawn in the following holds for any scenario and also when more complicated propagation models are used to determine G_{ij} parameters.

6.1. Performance in Regular Topology. We first look at the performance of the proposed approach in the regular topology, as illustrated in Figure 4(a). A total number of $n = 9$ nodes are placed in a 3×3 grid, and the deployment area is a square size of 80×80. The unit grid separation (i.e., distance between adjacent nodes along the grid-side) is 20 m. All nodes use a common transmit power of 3 mW, which results in a transmission range of 23.4 m. The transmission range is greater than the unit grid separation but is less than the unit grid diagonal (i.e., distance between adjacent nodes along the diagonal). This results in a simple topology where

Table 2: Set of available channels at each node (i.e., C_i) for regular topology.

Node index	Location	Available channels
1	(20, 20)	2, 3, 4
2	(20, 40)	1, 2, 4, 6
3	(20, 60)	1, 2, 5,
4	(40, 20)	3, 4, 5, 6
5	(40, 40)	1, 3, 4, 6
6	(40, 60)	1, 3, 5
7	(60, 20)	1, 2, 5, 6
8	(60, 40)	2, 4, 5
9	(60, 60)	1, 2, 3, 6

Table 3: Set of available channels at each node (i.e., C_i) for random topology.

Node index	Location	Available channels
1	(29.0, 28.8)	1, 2, 6, 7
2	(23.2, 9.9)	3, 4, 5, 7
3	(12.4, 13.7)	5, 7, 8
4	(4.4, 38.8)	3, 4, 6
5	(15.3, 2.9)	1, 2, 3, 4, 7, 8
6	(12.8, 30.5)	2, 5
7	(38.0, 10.8)	1, 4, 6
8	(1.6, 6.1)	3, 4, 5
9	(24.3, 37.7)	2, 4, 5, 8
10	(39.1, 34.1)	1, 3, 6, 8

all nodes can only communicate with their physical one-hop neighbors on the grid. Figure 4(b) shows the connection graph. There are $M = 6$ channels that can be used for the entire network. Every node has 3 radio interfaces (γ_i). The set of available channels at each node is randomly generated; see Table 2. Note that the set of available channels is different from node to node.

6.1.1. Complexity Reduction. For this scenario, there are 12 potential links and 24 binary variables (i.e., $\{x_e^m\}$). By using the enumerative technique, we obtain 22 independent sets and the size of the maximum independent set is 2. The exhaustive search space to determine the binary variable is 2^{24}; however, combining the independent sets together with the rounding and local optimization, the search space is significantly reduced from 2^{24} to 1800.

6.1.2. Throughput. Regarding the traffic flow, we consider $|Q| = 1$, 2, or 3 active sessions and run 3 experiments, respectively. In each experiment, the source node and destination node for each session are randomly generated. Figure 5(a) shows the results of the throughput obtained by our heuristic, upper bound and global optimum. Since the regular topology is simple, the heuristic method includes rounding technique only. It is observed that such heuristic results (obtained by rounding technique only) are equal to the global optimum, therefore no need to carry out local optimization. It is also found that there are gaps between the heuristic results and the relaxation bound.

For comparison purpose Figure 5(b) shows the results of the optimality ratio (which is defined as the normalized throughput over the global optimum) obtained by our rounding technique. It is found that the optimality ratio obtained by rounding is 1, while the optimality ratio of the relaxation bound is within (1,1.6). Simulation results show that the rounding technique performs very well in this scenario.

6.2. Performance in Random Topology. We next relax the regularity of node placement and look at the performance of the proposed approach in the random topology. As Figure 6(a) shows, we assume that $n = 10$ nodes are uniformly distributed in a square size of 40×40 area. All nodes use a common transmit power of 4 mW, which results in a transmission range of 25.1 m. Figure 6(b) shows the connection graph. There are $M = 8$ channels that can be used for the entire network. Every node has 4 radio interfaces (γ_i). The set of available channels at each node is shown in Table 3. Again, the set of available channels is different from node to node.

6.2.1. Complexity Reduction. For this scenario, there are 21 potential links and 32 binary variables (i.e., $\{x_e^m\}$). By using the enumerative technique, we get 29 independent sets and the size of the maximum independent set is 2. The exhaustive search space to determine the binary variable is 2^{32}; however, combining the independent sets together with the rounding and local optimization, the search space is significantly reduced from 2^{32} to 13500.

6.2.2. Throughput. Regarding the traffic flow, we consider $|Q| = 1$, 2, 3, 4, or 5 active sessions and run 15 experiments, respectively. In each experiment, the source node and destination node for each session are randomly generated. Different from the results obtained in the regular topology, in this random topology we show not only the heuristic results obtained by rounding technique but also the heuristic solutions obtained by rounding and local optimization. Table 4 shows the results. It is observed that there are some minor gaps between global optimum and the heuristic results obtained by rounding technique. However, by further using local optimization method, we find that the heuristic results are very close to the the global optimum. This observation demonstrates that the local optimization can lead to significant improvement on the objective function. Also note that there are some moderate gaps between the global optimum and the bounds obtained by relaxation.

For comparison purpose Table 5 shows the optimality ratio obtained by our heuristic and relaxation. It is observed that the heuristic results obtained by rounding and local optimization are very close to 1, while the heuristic results obtained by only rounding are within (0.48,1.0) and the ratio of the relaxation is within (1,1.8). The simulation results show that the combination of rounding and local optimization performs very well in this scenario.

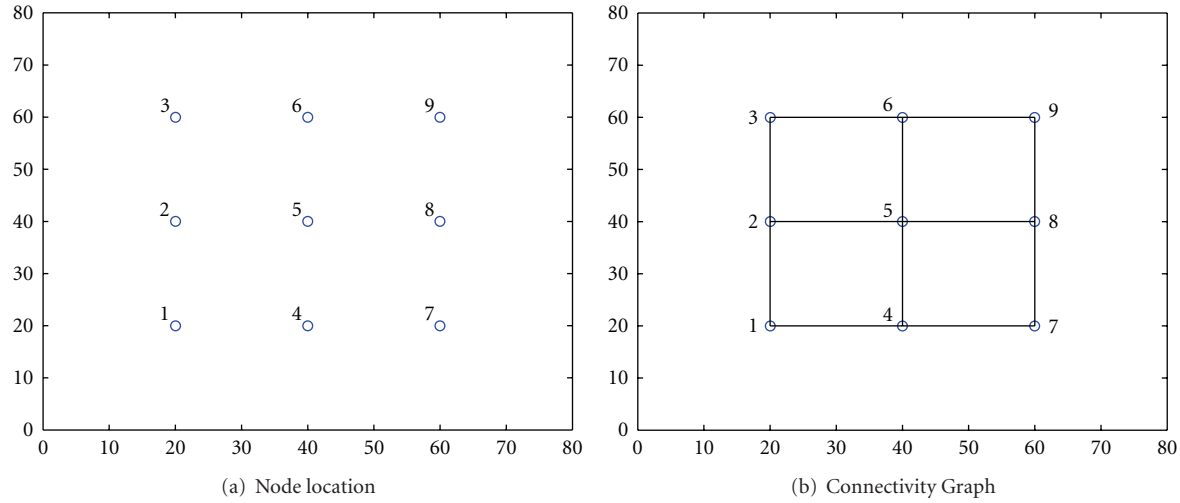


Figure 4: Regular topology.

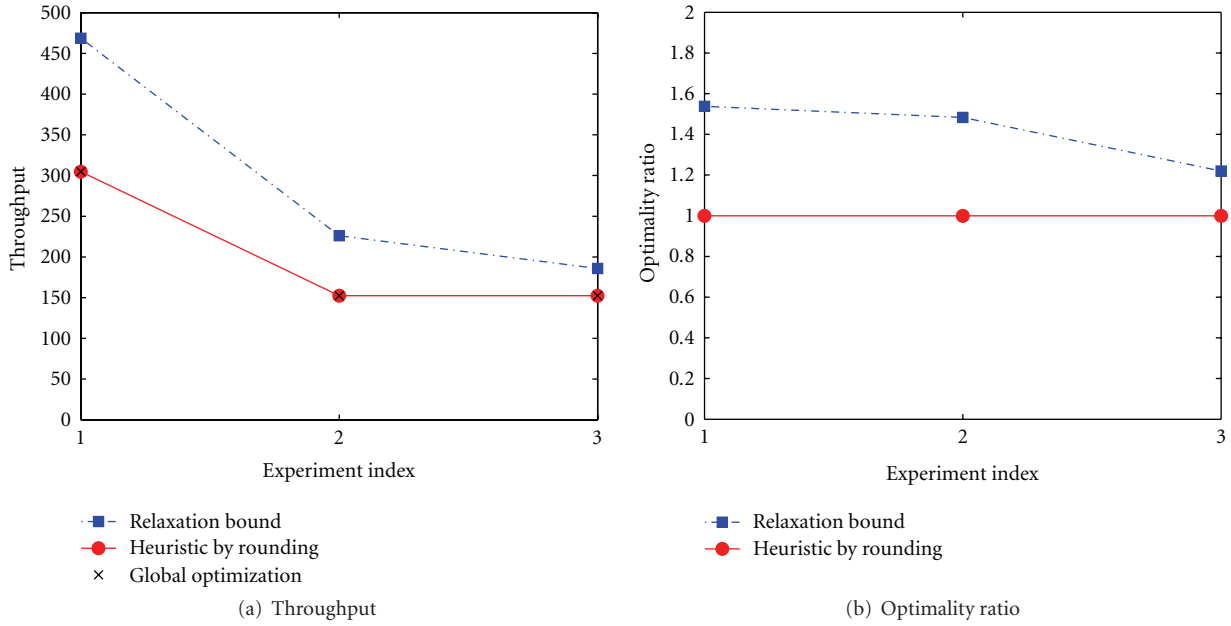
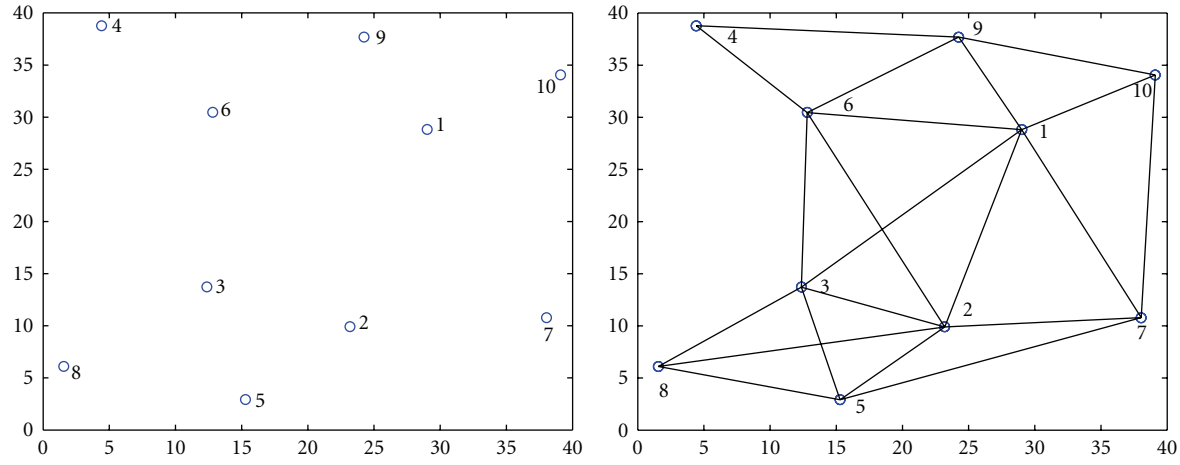


Figure 5: Comparison between heuristic, upper bound and global optimum for regular topology.

6.2.3. Spectrum Sharing and Routing. For illustration purpose, we show the results of spectrum sharing and routing when there are 5 communication sessions. The source node and destination node for each communication session are randomly generated; see Table 6.

By solving the MINLP problem by heuristic, we obtain that the optimal achievable throughput for each traffic flow is 111.3 (which is the 15th experiment shown in Figure 7). Figure 7(a) illustrates the optimal spectrum sharing. It is noticed that there are 10 active links in total, and channels 4 and 8 are reused. The nodes form themselves as an ad hoc network and all links can be active simultaneously (i.e., the topology is conflict-free). Figures 7(b)–7(f) illustrate the routing path(s) for each traffic flow. Figure 7(b) shows that the traffic flow generated by node 8 first travels to node 5, and then the traffic is split into 2 paths: one is 5 → 3 → 6 → 9 → 10 and the other path is 5 → 2 → 1 → 10. Figure 7(c) indicates that the traffic from node 7 to node 4 is via a single path, that is, 7 → 1 → 10 → 9 → 4. Similarly, as shown in Figures 7(d) and 7(e), the routing path for traffic flow from node 5 to node 9 is 5 → 3 → 6 → 9, while the traffic flow from node 3 to node 1 is via the path 3 → 5 → 2 → 1. Finally, the traffic flow generated from node 2 travels through 2 paths, one is through 2 → 5 → 3 → 6, and the other is via 2 → 1 → 10 → 9 → 6.



(a) Node location (b) Connectivity graph

Figure 6: Random topology.

Table 4: *Throughput* comparison for random topology.

Experiment index	Global optimal	Relaxation bound	Rounding	Rounding & local optimization
1	404.67	725.38	260.50	404.67
2	508.28	860.37	504.81	504.81
3	534.48	860.37	260.50	534.48
4	183.67	183.67	154.17	183.67
5	254.14	416.66	144.17	254.14
6	247.78	422.54	123.89	247.78
7	154.17	183.67	117.63	154.17
8	152.76	183.67	101.84	152.76
9	222.72	286.79	123.89	222.72
10	177.44	286.79	123.89	177.44
11	140.07	183.67	126.48	140.07
12	130.25	183.67	123.89	130.25
13	123.89	215.09	117.63	117.63
14	148.48	183.67	94.86	148.48
15	111.34	167.47	104.54	111.34

Table 5: *Optimality ratio* comparison for random topology.

Experiment index	Relaxation bound	Rounding	Rounding & local optimization
1	1.79	0.64	1.00
2	1.69	0.99	0.99
3	1.61	0.49	1.00
4	1.00	0.84	1.00
5	1.64	0.57	1.00
6	1.71	0.50	1.00
7	1.19	0.76	1.00
8	1.20	0.67	1.00
9	1.29	0.56	1.00
10	1.62	0.70	1.00
11	1.31	0.90	1.00
12	1.41	0.95	1.00
13	1.74	0.95	0.95
14	1.24	0.64	1.00
15	1.50	0.94	1.00

Table 6: Rate requirements of 5 sessions for random topology.

Source node	Destination node
8	10
7	4
5	9
3	1
2	6

7. Conclusion

In this paper, we consider a multihop multi-channel CR network. We present a cross-layer optimization framework by jointly designing the spectrum sharing and routing with the SINR constraints. Distinguished from the previous studies, we adopt a more realistic SINR model to capture the conflict relationships among the links, rather than using the Protocol Model. Our objective is to maximize the minimum end-to-end flow throughput, and our study addresses the following two cross-layer throughput optimization problem. (1) Given a set of secondary users with random but fixed location, and a set of traffic flows, what is the max-min achievable throughput? (2) To achieve the optimum, how to choose the set of active links, how to assign the channels to each active link, and how to route the flows? We answer these questions via a formal mathematical formulation in the forms of *mixed integer nonlinear programming (MINLP)*. Since the MINLP formulation is generally an NP-hard problem, we develop a heuristic method by solving a relaxation

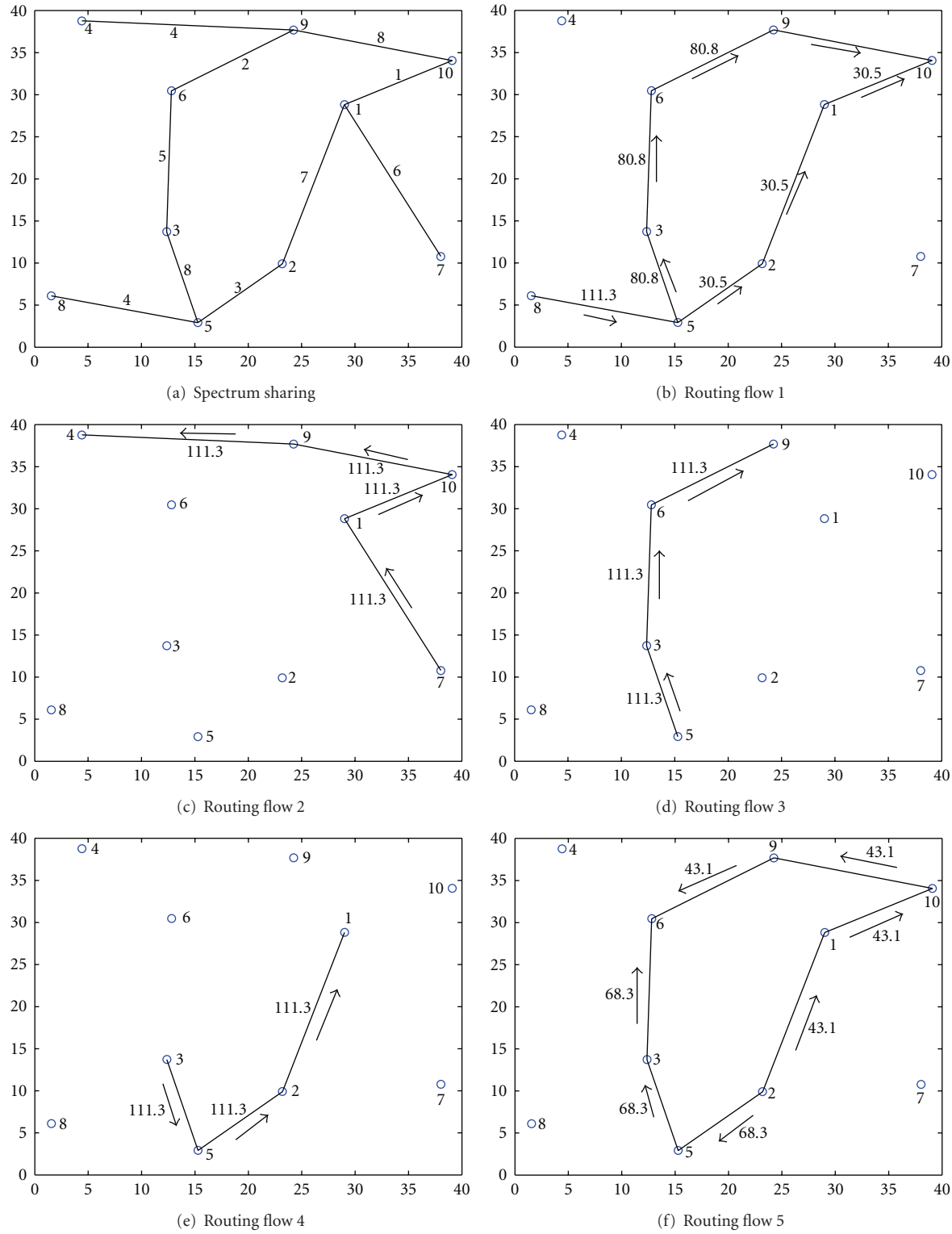


Figure 7: Spectrum sharing and routing for random topology.

of the original problem, followed by rounding and simple local optimization. Simulation results show that the heuristic approach performs very well; that is, the solutions obtained by the heuristic are very close to the global optimum.

For the future work, we need to consider how to design a distributed algorithm for a multihop CR network. Since in reality, there may not exist a centralized server, and also, the available channels are highly dynamic, in such

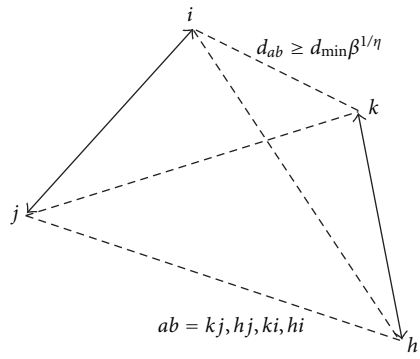


Figure 8: SINR model of interference.

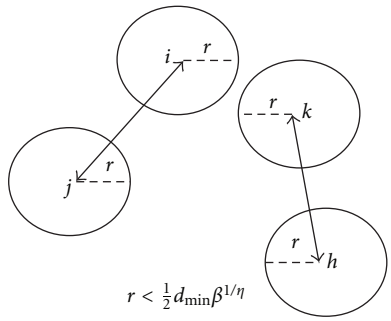


Figure 9: Disks of radius r centered at transmitters and receivers are *disjoint.*

situation, how to choose the set of active links and how to allocate channels and route the flows to obtain the max-min achievable throughput is a highly desirable and challenging work.

Appendix

Proof of Theorem 1

Consider two links, say $e = (i, j)$ and $e' = (k, h)$, using a same channel for simultaneous transmission. According to the SINR model, the transmission on links e is successful if and only if

$$\frac{P\left(d_0/d_{ij}\right)^{\eta}}{N + P(d_0/d_{ab})^{\eta}} \geq \beta, \quad \text{for } ab = kj, hj, ki, hi. \tag{A.1}$$

Then we find that d_{ab} is bounded by $d_{ij}\beta^{1/\eta}$, that is,

$$d_{ab} \geq d_{ij}\beta^{1/\eta}, \quad \text{for } ab = kj, hj, ki, hi. \tag{A.2}$$

Suppose that the minimum distance of any two nodes is $d_{\min}$; we have

$$d_{ab} \geq d_{\min}\beta^{1/\eta}, \quad \text{for } ab = kj, hj, ki, hi. \tag{A.3}$$

Thus, for links e and e' to be independent, the distances d_{kj}, d_{hj}, d_{ki}, and d_{hi} have to exceed $d_{\min}\beta^{1/\eta}$ (see Figure 8). This is equivalent to embed the disjoint disks of radius $d_{\min}\beta^{1/\eta}/2$ centered at both transmitters and receivers, as shown in Figure 9. Different from the case of the unidirectional link [16, 25], for the bidirectional link both the transmitter and the receiver "consume" a "wireless footprint" of area, and the total consumed area of a bidirectional link is $2\pi(d_{\min}\beta^{1/\eta}/2)^2$. Note that the total area of the domain is L^2 square meters. Thus, at most $2L^2/\pi(d_{\min}\beta^{1/\eta})^2$ transmissions are simultaneously feasible. That is, the number of simultaneous transmissions on a same channel is upper bounded by $2L^2/\pi(d_{\min}\beta^{1/\eta})^2$.

References

[1] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[4] Y. T. Hou, Y. Shi, and H. D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 1–9, May 2007.

[5] Y. Shi and Y. T. Hou, "Optimal power control for multi-hop software defined radio networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 1694–1702, May 2007.

[6] Y. Shi and Y. T. Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," in *Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM '08)*, pp. 1966–1974, April 2008.

[7] M. Ma and D. H. K. Tsang, "Efficient spectrum sharing and power control in cognitive radio networks," in *Proceedings of the 1st International Workshop on Cognitive Wireless Networks (CWNETS '07)*, 2007.

[8] M. Ma and D. H. K. Tsang, "Joint design of spectrum sharing and routing with channel heterogeneity in cognitive radio networks," *Physical Communication*, vol. 2, no. 1-2, pp. 127–137, 2009.

[9] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless netwrok performance," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MOBICOM '03)*, pp. 66–80, September 2003.

[10] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" in *Proceedings of the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine '06)*, vol. 191, Waterloo, Canada, August 2006.

[11] G. Brar, D. M. Blough, and P. Santi, "Computationally efficient scheduling with the physical interference model for throughput improvement in wireless mesh networks," in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MOBICOM '06)*, pp. 2–13, September 2006.

[12] D. Chafekar, V. S. A. Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan, "Cross-layer latency minimization in wireless networks with SINR constraints," in *Proceedings of the 8th*

ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07), pp. 110–119, September 2007.

[13] A. Behzad and I. Rubin, "Optimum integrated link scheduling and power control for multihop wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 194–205, 2007.

[14] LINDO Systems, *LINGO: User's Guide*, 2006.

[15] C. Xin, B. Xie, and C.-C. Shen, "A novel layered graph model for topology formation and routing in dynamic spectrum access networks," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 308–317, November 2005.

[16] S. Narayanaswamy, V. Kawadia, R. S. Sreenivas, and P. R. Kumar, "Power control in ad-hoc networks: theory, architecture, algorithm and implementation of the COMPOW protocol," in *Proceedings of European Wireless*, 2002.

[17] B. Hajek and G. Sasaki, "Link scheduling in polynomial time," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 910–917, 1988.

[18] G. Wang and N. Ansari, "Optimal broadcast scheduling in packet radio networks using mean field annealing," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 2, pp. 250–259, 1997.

[19] J. P. Monks, V. Bharghavan, and W. W. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proceedings of the 20th Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '01)*, pp. 219–228, April 2001.

[20] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad-hoc networks," in *Proceedings of the 21th Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '02)*, pp. 976–984, June 2002.

[21] A. Muqattash and M. Krunz, "Power controlled dual channel (PCDC) medium access protocol for wireless ad hoc networks," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '03)*, pp. 470–480, April 2003.

[22] Y. Xi and E. M. Yeh, "Distributed algorithms for spectrum allocation, power control, routing, and congestion control in wireless networks," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 180–189, September 2007.

[23] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[24] L. Badia, A. Erta, L. Lenzini, and M. Zorzi, "A general interference-aware framework for joint routing and link scheduling in wireless mesh networks," *IEEE Network*, vol. 22, no. 1, pp. 32–38, 2008.

[25] A. Karnik, A. Iyer, and R. C. Rosenberg, "Throughput-optimal configuration of fixed wireless networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1161–1174, 2008.

[26] A. Zymnis, S. Boyd, and D. Gorinevsky, "Relaxed maximum a posteriori fault identification," *Signal Processing*, vol. 89, no. 6, pp. 989–999, 2009.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 236568, 11 pages
doi:10.1155/2010/236568

Review Article

A Survey of Cognitive Radio Access to TV White Spaces

Maziar Nekovee[1, 2]

[1] *BT Innovate and Design, Polaris 134, Adastral Park, Martlesham, Suffolk IP5 3RE, UK*
[2] *Centre for Computational Science, University College London, 20 Gordon Street, London WC1H 0AJ, UK*

Correspondence should be addressed to Maziar Nekovee, maziar.nekovee@bt.com

Received 2 November 2009; Accepted 6 April 2010

Academic Editor: Fred Daneshgaran

Cognitive radio is being intensively researched as the enabling technology for license-exempt access to the so-called TV White Spaces (TVWS), large portions of spectrum in the UHF/VHF bands which become available on a geographical basis after digital switchover. Both in the US, and more recently, in the UK the regulators have given conditional endorsement to this new mode of access. This paper reviews the state-of-the-art in technology, regulation, and standardisation of cognitive access to TVWS. It examines the spectrum opportunity and commercial use cases associated with this form of secondary access.

1. Introduction

A cognitive radio [1] consists of a cognitive engine (CE), which contains algorithms and toolboxes for radio environment sensing, machine-learning, and reasoning and decision making, and a configurable radio platform, which could be a Software Defined Radio (SDR), that basically does what it is told by the CE. The concept of Cognitive Radio (CR) was first described by Mitola and Maguire [2] as "transforming radio nodes from blind executors of predefined protocols to radio-domain-aware intelligent agents that search out ways to deliver the services that the user wants even if that user does not know how to obtain them". The ideal CR knows everything about the user requirements, the capability of the radio device, the network requirements and the external environment (including the radio environment). It will plan ahead and negotiate for the best part of the spectrum to operate in and at the best power, modulation scheme, and so forth, and manage these resources in real time to satisfy the service and user demands. The ideal CR is currently at the early proof-of-concept stage research, with most of the work taking place in universities.

A much more developed form of the CR technology is cognitive radio for dynamic spectrum access (DSA) [3]. The aim here is to achieve device-centric interference control and dynamic reuse of radio spectrum based on the frequency agility and intelligence offered by cognitive radio technology. This form of CR technology is currently being intensely researched. However, there is also already significant industry effort towards prototyping, standardisation and commercialisation of the technology. Important industry players with active R&D efforts in cognitive radio technology include Alcatel-Lucent, Ericsson and Motorola from the mobile equipment industry, BT and Orange from network operators, Philips and Samsung from the consumer electronics industry, HP and Dell from the computer industry, and Microsoft and Google from the Internet/software industry. Dynamic spectrum access may take place in several ways: between a licensed primary system and a license-exempt secondary system, for example, secondary spectrum access to digital TV or military spectrum, within the same primary system, for example, micro-macro sharing of licensed spectrum in 3G/LTE femtocells, and finally among two primary systems, for example, real-time leasing and trading of spectrum between two cellular operators.

The first form of dynamic spectrum access is arguably the most disruptive application of the CR technology, as it enables license-exempt users (end-user devices and base stations) to act as spectrum scavengers. They can identify unused portions of licensed spectrum (also called spectrum holes or White Spaces) and make opportunistic use of this spectrum for their connectivity at times and/or locations where they are not used. Allowing the operation of such scavengers promises to greatly increase the efficiency of

spectrum usage by preventing exclusively licensed spectrum from being wasted due to low spatial or temporal usage. Mainly for this reason licensed-exempt cognitive access to certain licensed bands is being keenly promoted by the US regulator, the (Federal Communication Commission) FCC [4–7], and more recently also by Ofcom [8–10]. The rationale is to maximise the usage of licensed spectrum through secondary access by cognitive radios and, at the same time, promote rapid introduction of new wireless technologies and services without the need for setting aside any new spectrum for this purpose. Most mobile operators see this from of cognitive access as highly disruptive to their current business model.

In the longer term (3–5 years), we expect that dynamic spectrum access based on cognitive radio will go far beyond opportunistic spectrum access only. As a result of the current trends in spectrum liberalisation, including the availability of licensed spectrum for real-time trading, cognitive devices may be able to access a portfolio of different types of spectrum for their connectivity. This "spectrum portfolio" may include several different type of spectrum: licensed spectrum (e.g., in cellular bands), licensed-exempt spectrum (in the ISM bands), as well as spectrum, that is, acquired in real-time, either through leasing or on a secondary basis. Devices with cognitive functionality will be able to dynamically change their operating spectrum within this portfolio, accessing the best available spectrum on a "just-in-time" basis. This may happen either upon instruction from a base station or autonomously by devices themselves. Depending on the user and network requirements devices may pool together and use several spectrum fragments and vacate some or all of them when they are no longer required or when other more suitable ones become available. These requirements may depend on context, application and location and can include price, Quality of Service (QoS), and energy saving.

To date both in the UK [10] and US [5–7] regulators have committed to licence-exempt cognitive access to the so-called TV White Spaces (TVWS). The TVWS spectrum comprises large portions of the UHF/VHF spectrum that become available on a geographical basis for cognitive access as a result of the switchover from analogue to digital TV. The total capacity associated with TVWS is significant. According to modelling studies commissioned by Ofcom over 50% of locations in the UK are likely to have more that 150 MHz of interleaved spectrum and that even at 90% of locations around 100 MHz of interleaved spectrum might be available for cognitive access [10]. In addition to TVWS, the defence spectrum may provide another significant capacity opportunity for license-exempt cognitive access. For example, around 30% of spectrum below 15 GHz is allocated to Defence in the UK. The UK (Ministry of Defence) MoD had until the late 1990's access to spectrum at no or a low cost. However, following the Cave Audit, the Government has committed to releasing a "significant proportion" of the MOD's spectrum between 2008 and 2010. Results form a 2008 study by PA consulting (commissioned jointly by MoD and Ofcom) suggest that [11] there is significant scope for license-exempt use of the released spectrum using cognitive radio technology, both on a spatial and a temporal basis. For example, low power cognitive devices could potentially share with radar if the radar sweep can be detected and the transmission of the cognitive device can be timed to avoid interference.

This paper aims to review the state-of-the-art in technology, regulation, and standardisation of cognitive radio access to TVWS. It also examines the spectrum opportunity, potential business applications, and some of the open research challenges associated with this new form of access, drawing lessons and conclusions from recent recent findings in the UK [12, 13], US [14, 15], and elsewhere. The rest of this paper is organised as follows. Section 2 provides a brief overview of cognitive radio access to TV White Spaces. In Section 3 the regulatory status and standardisation efforts are reviewed and some of the outstanding research and technology challenges are discussed. In Section 4 we discuss recent results on quantifying the availability of TVWS spectrum for cognitive access in the UK and the US, and describe some of the prominent candidate use cases of this spectrum. We conclude this paper in Section 5.

2. Cognitive Access to TV White Spaces

2.1. What are TV White Spaces? Broadcast television services operate in licensed channels in the VHF and UHF portions of the radio spectrum. The regulatory rules in most countries prohibit the use of unlicensed devices in TV bands, with the exception of remote control, medical telemetry devices, and wireless microphones. In most developed countries regulators are currently in the process of requiring TV stations to convert from analogue to digital transmission. This Digital Switchover (DSO) was completed in the US in June 2009, and is expected to be completed in the UK by 2012. A similar switchover process is also underway or being planned (or is already completed) in the rest of the EU and many other countries around the world. After Digital Switchover a portion of TV analogue channels become entirely vacant due to the higher spectrum efficiency of digital TV (DTV). These cleared channels will then be reallocated by regulators to other services through auctions.

In addition to cleared spectrum, after the DTV transition there will be typically a number of TV channels in a given geographic area that are not being used by DTV stations, because such stations would not be able to operate without causing interference to cochannel or adjacent channel stations. However, a transmitter operating on such a locally vacant TV channel at a much lower power level would not need a great (physical) separation from cochannel and adjacent channel TV stations to avoid causing interference. Low power devices can therefore operate on vacant channels in locations that could not be used by TV stations due to interference planning. These vacant TV channels are known as TV White Spaces or Interleaved Spectrum in the language of the UK regulator.

2.2. Detection and Incumbent Protection. Secondary operation of cognitive radios in TV bands relies on the ability of cognitive devices to successfully detect TVWS, and is

conditioned by regulators on the ability of these devices to avoid harmful interference to licensed users of these bands, which in addition to DTV include also wireless microphones. Both the FCC and Ofcom have considered three methods for ensuring that cognitive devices do not cause harmful interference to incumbent: beacons, geolocation combined with access to a database, and sensing. Currently, the database approach seems to offers the best short-term solution for incumbent detection and interference avoidance. Both in the US and UK regulatory and industry efforts is, therefore, underway to further develop the concepts, algorithms and regulatory framework necessary for this approach.

2.2.1. Beacons. With the beacon method, unlicensed devices only transmit if they receive a control signal (beacon) identifying vacant channels within their service areas. The signal can be received from a TV station, FM broadcast station, or TV band fixed unlicensed transmitter. Without reception of this control signal, no transmissions are permitted. One issue with the control signal method is that it requires a beacon infrastructure to be in place, which needs to be maintained and operated, either by the incumbent or a third party. Furthermore, beacon signals can be lost due to mechanisms similar to the hidden node problem described below.

2.2.2. Geolocation Combined with Database. In this method, a device determines its location and accesses a database to determine the TV channels that are vacant at that location. There are at least three issues associated with this method. There is a need for a new (commercial) entity to build and maintain the database. Devices need to know their location with a prescribed accuracy. For outdoor applications GPS can be used to support these requirements, but in the case of indoor application there are issues with the penetration of GPS deep. Finally, devices need additional connectivity in a different band in order to be able to access the database prior to any transmission in DTV bands. inside buildings.

We note that the latter problems can be addressed in master-slave communication architectures where a master device, for example, an access point or base station, has access to location information and is connected via a wireless or fix link to the Internet. The master node uses its location information to query the geolocation database about TVWS channel availability and based on this information instructs a set of slave devices on the frequencies they can use.

2.2.3. Sensing. Finally, in the sensing method, unlicensed devices autonomously detect the presence of TV signals and only use the channels that are not used by TV broadcaster. Detection of the TV signal can be subject to the hidden node problem, which is depicted in Figure 1. This problem can arise when there is blockage between the unlicensed device and a TV station, but no blockage between the TV station and a TV receiver antenna and no blockage between the unlicensed device and the same TV receiver antenna. In such a case, a cognitive radio may not detect the presence of the TV signal and could start using an occupied channel, causing harmful interference to the TV receiver.

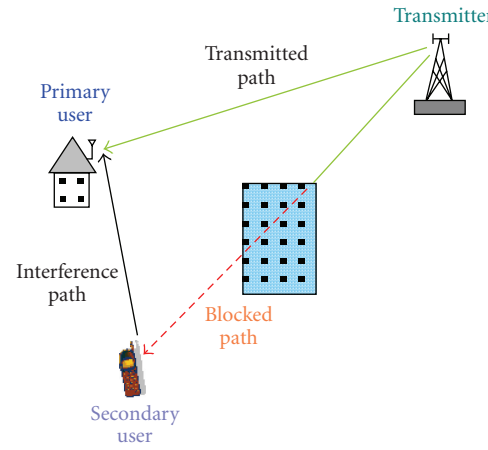


Figure 1: The hidden node problem of a sensing-based cognitive radio.

2.3. Regulatory Developments

2.3.1. US. In the US the FCC proposed to allow opportunistic access to TV bands already in 2004 [4]. Prototype cognitive radios operating in this mode were put forward to FCC by Adaptrum, (Institute for Infocomm Research) I2R, Microsoft, Motorola, and Philips in 2008. After extensive tests the FCC adopted in November 2008 a Second Report and Order that establishes rules to allow the operation of cognitive devices in TVWS on a license-exempt basis [5–7]. In summary these rules require cognitive devices to use both spectrum sensing and geolocation. In order to minimise the chance of harmful interference due to the hidden node problem FCC has required that cognitive devices should be able to sense both television signals and wireless microphones down to −114 dBm. They must also locate their position to within 50 metres and then consult a database that will inform them about available spectrum in that location [5–7].

Mobile devices may transmit in a locally vacant TV channel at up to 100 mW unless they are using a channel adjacent to terrestrial television, in which case their transmission power can only be 40 mW. Fixed devices (base stations or customer premises) may transmit at a locally vacant channel at up to 4 W (EIRP). Devices without geolocation capabilities are also allowed if they are transmitting to a device that has determined its location. In this case, one device would be acting as a master for a network and the other slave devices would operate broadly under its control in terms of the spectrum they would use. Devices that use sensing alone are allowed in principle; however, they must be submitted in advance to the FCC for laboratory and field testing so the FCC can determine whether they are likely to cause harmful interference. The exact process that the FCC will use to determine this has not been specified.

Importantly, the FCC report includes a detailed discussion about whether cognitive access should be licensed, licence-exempt or subject to light licensing. It concludes that the best way to facilitate innovative new applications is via licence-exemption and that licensing would not be

practicable for many of the new applications envisaged. It also notes that any licenses would be difficult to define and subject to change (e.g., if television coverage was replanned), so the rights awarded would be rather tenuous.

2.3.2. UK. In its Digital Dividend Review Statement released in December 2007 the UK regulator, Ofcom, proposed to "allow licence exempt use of interleaved spectrum for cognitive devices" [8]. Furthermore Ofcom stated that "We see significant scope for cognitive equipments using interleaved spectrum to emerge and to benefit from international economics of scale [8]". In a consultation published on 16 February 2009 [9] Ofcom proposed a number of technical parameters for licence-exempt cognitive use of interleaved spectrum which closely follow those suggested by FCC.

Subsequently, in a statement published on July 1 2009 Ofcom proposed to allow sensing alone as well as geolocation for incumbent detection [10]. However, it concludes that in the short term the most important mechanism for spectrum detection will be geolocation. Ofcom is suggesting that further work, possibly leading to a consultation specifically on geolocation, is appropriate. Finally Ofcom states that it "will work with stakeholders to further develop the concepts and algorithms necessary for geolocation and expect to consult further on Geo-location later in 2009" [10]. Following this consultation Ofcom published a discussion paper on geolocation for cognitive access in Novemeber 2009.

2.3.3. Worldwide. Work on a pan-European specification for cognitive devices is currently taking place within the SE43 working group of (the European Conference of Postal and Telecommunications Administrations) CEPT. An important aim of this group is to define technical and operational requirements for the operation of cognitive radio systems in TV White Spaces in order to ensure the protection of incumbent services/systems and to investigate the amount of spectrum across Europe that is potentially available as White Spaces. Furthermore on a worldwide scale, agenda item 1.19 of the (World Radiocommunications Conference, 2011) WRC-11 will be considering regulatory measures and their relevance, in order to enable the introduction of software defined radio and cognitive radio systems, based on the results of ongoing ITU-R studies.

2.4. Standardisation and Industry Effort. Industry led research and development on cognitive radio technology has been so far mainly focused in the USA, and is largely driven by the desire of important new players, including Google and Microsoft, to get access to the TVWS spectrum. However, a number of EU-backed and industry-led collaborative projects are currently underway that aim at bringing cognitive radio technology in Europe closer to commercial exploitation. Two major standardisation efforts, which are currently at an advanced stage, are discussed below. It is worth mentioning that in addition to these a number of new standardisation initiatives are underway, which include the IEEE 802.19 and the IEEE 802.11af standards.

2.4.1. The Cognitive Networking Alliance (CogNeA) Standard. The Cognitive Networking Alliance (CogNea) [16] is an open industry association. The Alliance intends to commercialise low power personal/portable Cognitive Radio platforms by enabling and promoting the rapid adoption, regulation, standardisation and multivendor compliance and interoperability of CRs world wide. Alliance board members include ETRI, HP, Philips, Samsung Electro-Mechanics, Texas Instruments, and more recently BT. The initial geographical focus area is North America. The initial focus radio spectrum is TV White Spaces.

The Alliance intends to promote TVWS spectrum regulations worldwide, and to establish a recognisable CogNeA brand that indicates a device is CogNeA-compliant and can therefore interoperate with other CogNeA-certified devices from various manufacturers. The Alliance also develops specifications for the Common Cognitive Radio Platform (CCRP) which supports multiple applications [17].

The Alliance intends to bring the standard to an international status, in collaboration with an existing Standards Definition Organisation (SDO), to make it globally accepted. The primary target applications for the CogNeA standard are

(i) in-home high definition multimedia networking and distribution solutions that overcome the whole home coverage problems inherent to solutions using ISM bands,

(ii) unlicensed broadband wireless access for communities/neighbourhoods/campuses.

The standard is developing a Common Radio Platform consisting of the Physical Layer (PHY) and the Media Access Layer (MAC). The PHY consists of the Radio Front End, the Baseband, and the Cognitive Entity, which contains a geo-location block, a sensing block and an Internet access and interference map resources. The MAC carries the Communication/Networking protocol, Air access rules, and interface for the higher layers, such as network and application layers.

ECMA International is currently developing a high-speed wireless networking standard for use in the Television White Spaces, based on the contribution from CogNeA. The standard will employ cognitive radio sensing and database technologies to avoid interference with licensed services and other incumbent users in compliance with the FCC regulatory rules. The first draft of the ECMA/CogNeA PHY and MAC standard for operation in TV White Spaces was published in December 2009 [18, 19].

2.4.2. The IEEE 802.22 Standards. The IEEE 802.22 Working Group [20] has defined an air interface (PHY and MAC) standard based on cognitive radio techniques. The 802.22 standard is being developed for Wireless Regional Area Networks (WRANs). The primary target application of the standard is licensed-exempt broadband wireless access to rural areas in TVWS. The initial geographical focus area is North America. The 802.22 system specifies a fixed point-to-multipoint wireless air interface whereby a base station (BS) manages its own cell and all associated Consumer Premise

Equipments (CPEs). The network architecture including MAC and PHY are derived from IEEE 802.16 WiMAX. The 802.22 PHY layer is designed to support a system which uses vacant TVWS channels to provide wireless communication access over distances of up to 100 Km. The PHY specification is based on Orthogonal Frequency Division Multiple Access (OFDMA) for both the upstream and downstream access.

The IEEE 802.22 standard supports incumbent detection through spectrum sensing (the database approach is optional). The standard specifies inputs and outputs for the sensing function, as well as the performance requirements for the sensing algorithms implemented (e.g., probability of detection, incumbent detection threshold and probability of false alarms). These include energy detector, and cyclostationary and pilot sensing detectors for ATSC DTV signals, and an FFT-based algorithm for detection of wireless microphone signals [21].

The IEEE 802.22 defines a connection oriented and centralised MAC layer. Two important capabilities are introduced in the 802.22 MAC layer to support reliable incumbent detection based on sensing: network-wide quite periods scheduled by each BS during which all transmissions are suspended in order to allow reliable sensing, and channel measurement management to coordinate distributed channel measurement/incumbent detection by CPEs and their reporting to BS.

After an initial accelerated phase, the development of the standard seems to have slowed down during the last year. According to the IEEE 802.22 sources, the standard is currently at the Ballot stage. However, the final completion date for standard is not known yet, and there have been no vendor companies so far to build equipment based on the IEEE 802.22.

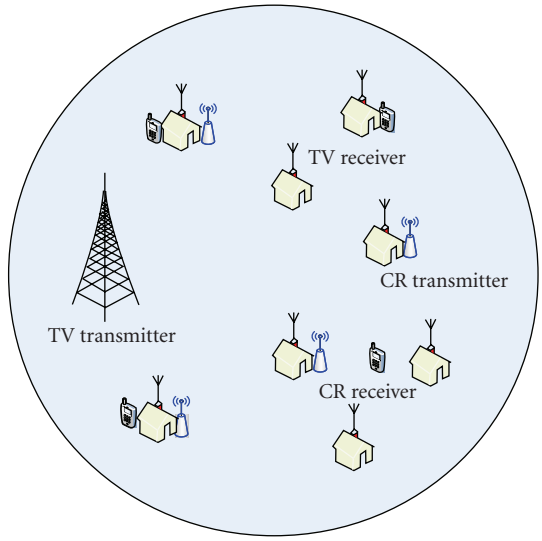


Figure 2: In future commercial applications the aggregate power levels of cognitive devices need to be controlled in order to avoid interference to primary receivers.

3. Research Challenges

3.1. High-Precision Spectrum Sensing. In order to minimise the chance of harmful interference due to the hidden node problem both the FCC and Ofcom require that cognitive devices should be able to sense TV signals at detection margins much lower than that of TV receivers (114 dBm for 6 MHz US channels and −120 dBm for 8 MHz UK channels) [5–7, 10]. Such weak signals are well below thermal noise, and in the presence of noise uncertainty cannot be detected using the energy detection algorithms that are implemented in the current generation of wireless devices [22]. Recent research, however, shows that such sensing levels may be achieved using more sophisticated sensing algorithms that rely on certain features of incumbent signals which are absent in the noise [23].

In the following we briefly discuss some of these "non-blind" sensing techniques. We note that, unlike energy detection, these algorithms are generally not applicable to sensing signals from wireless microphones, most of which use analogue frequency modulation (FM), and refer the reader to [24, 25] for recently proposed algorithms for detection of wireless microphones.

(i) *Pilot detection*. These sensing algorithms are specific to ATSC signals, which have a DC pilot at a lower band-edge in a known location relative to the signal. Detection is achieved by setting a threshold either on the amplitude or the location of the pilot signal. Detection based on the location of pilot is in particular robust against noise uncertainty, since the position of the pilot can be pinpointed with high accuracy, even if the amplitude is low due to fading [21].

(ii) *Cyclostationary feature detection*. Both the ATSC and DVB-T signals are cyclostationary, that is, the means and correlation sequences of these signals exhibit periodicity. Cyclostationary feature detectors were introduced as a complex two-dimensional signal processing technique for recognition of modulated signals in the presence of noise and interference. Recently they have been proposed by a number of authors [26, 27] for the detection of weak TV signals in the context of spectrum sensing for cognitive radio.

(iii) *Cyclic prefix and autocorrelation detection*. OFDM signals, including DVB-T signals, contain a special sequence called cyclic prefix (CP), where the last D bits of the OFDM symbol is copied to the beginning of the symbol. Cyclic prefix detection is similar to energy detection. However, the test statistics used in the algorithm is the energy contained in the cyclic prefix of each OFDM symbol, instead of the whole symbol [28]. Furthermore, due to the presence of CP the autocorrelation function of DVB-T signals show distinct peaks at nonzero values whose amplitude and position could be used to detect the signal from the noise.

One issue with most of the above sensing algorithms is that they require considerable processing power which may be either not available or not desirable (due to power consumption) in handheld devices. A second problem with high-precision sensing is that the ability of cognitive devices to sense extremely weak TV signals may eliminate the hidden node problem (false negatives) but at the same time it can lead to a situation where a cognitive radio detects TV signals from transmitters that are perhaps hundreds of kilometres away (false positive), thereby removing a considerable portion of usable White Spaces. Very recent studies in the US, for example, indicate that a threshold of −114 dBm reduces the recoverable White Spaces by a factor of 3 [14]. Even worse, initial modelling studies performed at BT [29] show that in some UK locations, a cognitive device with a −114 dBm sensitivity level will identify all DTT channels as occupied, and therefore will have no White Space available for its operation if it relies on naive sensing only!

There has been considerable recent research in cooperative detection algorithms, where sensing measurements performed by multiple devices are combined (using either a soft or a hard decision combining method) in order to achieve higher sensing thresholds than is possible by single devices or to deal with the hidden node problem [30–32]. Interestingly, cooperative sensing was also considered in Ofcom's consultation on cognitive access as a possible approach to the detection problem [9].

One problem of cooperative sensing is that the achievable detection level depends on several factors, including the number of cooperating cognitive devices and their spatial arrangement [30, 31]. Therefore, in general it would be difficult to test and certify the detection capability of such cooperating cognitive devices on an individual basis to check device compliance with regulatory requirements [29]. Furthermore this method requires additional communication overhead since local measurements will be collected at system level in order to make a decision, which is then broadcast to all cognitive radios involved.

Due to the above issues we believe that the most promising application of cooperative sensing will be in master-slave communication scenarios, where a computationally powerful master device (e.g., a WiFi access point or a cellular base station) centrally coordinates and process sensing activities of a set of slave devices in combination and uses the result to refine and geographically extend the results of its own sophisticated sensing algorithms.

3.2. Agile Transmissions and Spectrum Pooling Techniques. Physical layer transmission techniques that are able to effectively deal with the fragmented nature of TVWS spectrum are a very important component of future cognitive radios. In particular, these techniques must be sufficiently agile to enable unlicensed users to transmit in (locally) available TVWS bands while not interfering with the incumbent users operating at adjacent bands. Moreover, to support throughput-intensive applications, these techniques should be able to achieve high data rates by pooling several (not necessarily contiguous TVWS channels). One technique that seems to meet both these requirements is a variant of orthogonal frequency division multiplexing (OFDM) called *non-contiguous OFDM* (NC-OFDM) [33, 34]. NC-OFDM is capable of deactivating subcarriers across its transmission bandwidth that could potentially interfere with the transmission of other users. Moreover, NC-OFDM can support a high aggregate data rate with the remaining subcarriers, and simultaneously maintain an acceptable level of error robustness. In addition to NC-OFDM several other techniques have been proposed to enable agile waveforming over fragmented spectrum. One prominent example is the use of filterbank multicarrier techniques for such cognitive radio applications [35].

3.3. Multiple Antenna Technologies for Cognitive Radio. The use of antenna diversity or MIMO antenna architecture can provide a significant increase in the spectral efficiency of wireless systems [36, 37]. However, the use of multiple antennas in cognitive radio networks is underdeveloped. One of the major objectives for cognitive radio is to improve the spectrum utilisation. With the advantages offered by MIMO systems, it is therefore logical to exploit potentials in applying the MIMO antenna architecture to cognitive radio networks. Introducing multiple antenna technologies for cognitive radio (CR) may extend the dimension of CR from the current frequency band and time slot regime even further into spatial domain. A cognitive system using MIMO can significantly improve receiver sensitivity and coverage, hence it may also have impact on the key device parameters such as sensitivity and transmit power in CR, that is, required by regulators.

One issue with the use of MIMO in the context of cognitive access to TVWS is that the typical wavelengths in the UHF bands vary between 0.3–0.6 m. Optimal use of multiple antennas on a single cognitive device, therefore may not be feasible in most applications due to the small footprints involved. However, fixed BWA(Broadband Wireless Access) applications similar to that considered in the IEEE 802.22 that involve large base stations and customer premises may greatly benefit from multiple antenna technologies.

3.4. System-Level Issues. Almost all research so far has focused on a single cognitive device accessing TVWS spectrum. However, the provision of commercial services based on cognitive radio technologies, for example, mobile broadband or wireless home networks, will inevitably involve situations involving multiple cognitive equipments that may belong to either the same or different service providers. Some open research challenges associated with such service scenarios include the following:

(i) estimation and control of aggregate interference from multiple cognitive devices [38] towards primary users,

(ii) politeness (etiquette) rules that achieve fair and efficient sharing of secondary spectrum among competing cognitive radios [9],

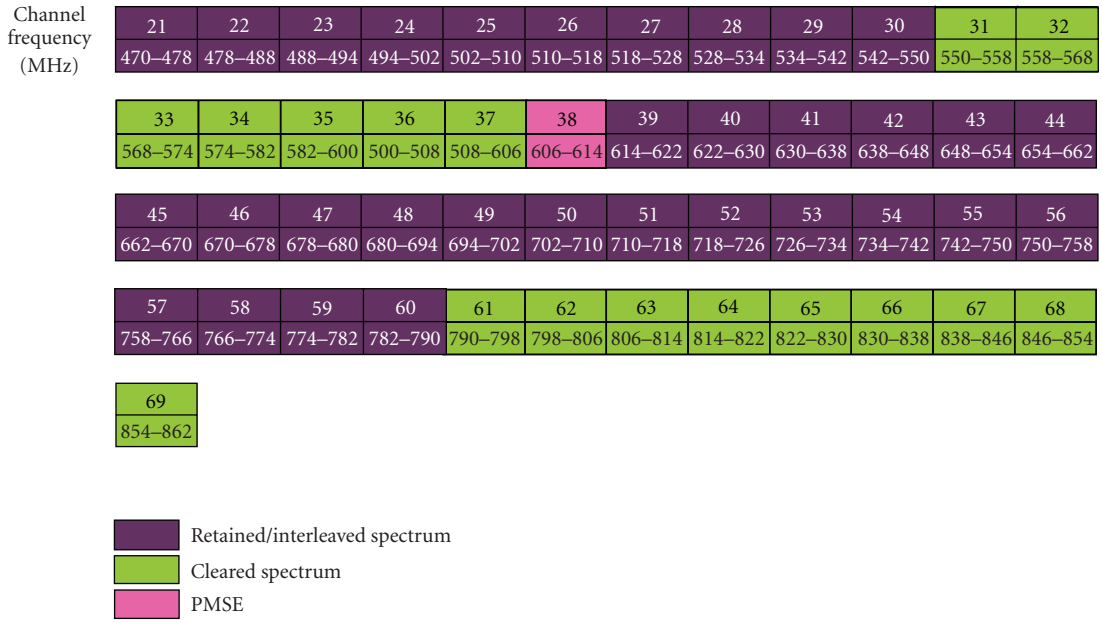


Figure 3: The UK UHF TV bands after completion of Digital Switchover [9].

(iii) spectrum sensing under aggregate interference conditions,

(iv) quantitative understanding of the scalability of secondary spectrum access networks operating in TVWS spectrum.

4. TVWS Spectrum Availability and Use Cases

4.1. How Much White Spaces Is Available? Figure 3 shows allocation of the UHF spectrum in the UK after the completion of DSO [10]. The 128 MHz of spectrum marked in green (16 bands) is the cleared spectrum which Ofcom plans to license through auctions. The 256 MHz (32 channels) marked in purple is the interleaved spectrum which can be used on a geographical basis for license-exempt access by using cognitive radio technology. Finally the channel marked in pink is licensed by Ofcom for exclusive access for wireless microphones, and so forth, (PMSE).

From the above chart it appears that there is significant capacity available for cognitive access in the UHF bands. However, due to its secondary nature the availability and frequency decomposition of the UHF spectrum for cognitive access is not the same at all locations and depends also on the power levels used by cognitive devices [12, 14]. This is an important feature of license-exempt cognitive access to TV bands which distinguish it from, for example, WiFi access to the ISM bands.

Potential commercial applications of TVWS devices will strongly depend on how the availability of this spectrum varies; both from location to location and as a function of transmit power of cognitive devices. A number of recent studies have investigated various aspect of TVWS spectrum in the US [14, 15]. In the UK we have developed a set of modelling tools that have enabled us to quantify the availability of TVWS spectrum for cognitive access and its variation with location and transmit power.

The first set of these modelling tools [12] makes use of the publicly available maps of DTV coverage in the UK [39] which were generated via computer simulations from the Ofcoms database of location, transmit power, antenna height and transmit frequency of UKs DTV transmitters, and were further validated and improved through direct observations at different locations. It combines these coverage maps with simplified propagation modelling calculations to obtain upper bounds for the vacant TVWS frequencies at any given location as well as a lower-bound estimate for the variations of TVWS spectrum with the transmit power of a cognitive devices.

The computer model for obtaining the upper bounds works as follows [12]. We use the UK National Grid (NG) coordinate system in order to specify the geographical position of any location on the UK map. Given the NG coordinates of a UK location the computer code then maps this location onto the closest grid point on the DTV coverage maps. For a given DTV transmitter this grid point is then evaluated to determine if it falls within the coverage area of that transmitter. If this is the case, then the frequencies associated with the transmitter are tagged as occupied at those locations, otherwise they are tagged as vacant. Repeating this procedure for coverage maps of all DTV transmitters, we then obtain a list of vacant TV frequencies at a given location that can be used by a low-power cognitive devices which is positioned in that location.

In the case of high power cognitive equipments, for example, those considered within the 802.22 standard, the required computations are very intensive. In order to reduce this computational effort, we approximate the actual DTV

coverage areas by circular disks which were constructed such that each of them entirely encompassed the coverage area of the associated transmitter while also having the minimum possible surface area. With this simplification, it is then computationally straightforward to calculate from the vacant TV frequencies as a function of both position and transmit power of cognitive devices.

A second modelling tool is currently in its final development stage [40]. The tool makes use of Ofcom's published database of DTV transmitter together with highly accurate terrain data and standard UHF radio propagation models to generate contour maps of the received power for every DTV transmitters in the UK. Combining these contour maps with propagation modelling of cognitive devices, the available TVWS for cognitive access at any given location and for any arbitrary power level is then computed with a spatial resolution of 100 m. The terrain data used in this modelling tool is based on the STRM v2 terrain elevation data. The STRM data set resulted from a collaboration effort between NASA and the US National Geographic Intelligence Agency, as well as the participation of the German and the Italian space agencies, and is at present the most complete high-resolution digital topographic database of Earth.

We have used our first set of modelling tools to investigate the variations in TVWS as a function of the location and transmit power of cognitive radios, and to examine how constraints on adjacent channel emissions of cognitive radios may affects the results. This analysis provides a realistic view on the potential spectrum opportunity associated with cognitive radio access to TVWS in the UK, and also presents the first quantitative study of the availability and frequency composition of TVWS outside the United States. Figure 4 summarises in a bar-chart the availability of TVWS channels for 18 major population centres in England, Wales and Scotland. The total number of channels available at each location is shown as a green bar. These results show that there are considerable variations in the number of TVWS channels as we move from one UK location to another. For any given location, however, a minimum of 12 channels (96 MHz) is accessible to low-power cognitive devices, while the averaged per location capacity is just over 150 MHz.

When a high power cognitive device operates in a vacant TV channel, energy leakage to adjacent channels may cause interference to adjacent frequencies, which may be occupied. Ofcom had raised concerns that operation of low-power cognitive devices on a given channel may also cause adjacent-channel interference for mobile TV receivers that are in close vicinity. Consequently, even in some future use cases, cognitive devices may be constrained not to use vacant channels whose immediate adjacent frequencies are used for mobile TV. The total number of available TVWS after imposing the above adjacent channel constraint are shown as red bars in Figure 4. It can be seen that imposing the constraint greatly reduces the amount of accessible spectrum in most locations considered (on average the available capacity drops to just below 40 MHz/location).

Recent studies on quantifying the availability of TVWS in the United States were reported in [14, 15], and the results are in line with our findings for the UK. In particular a detailed

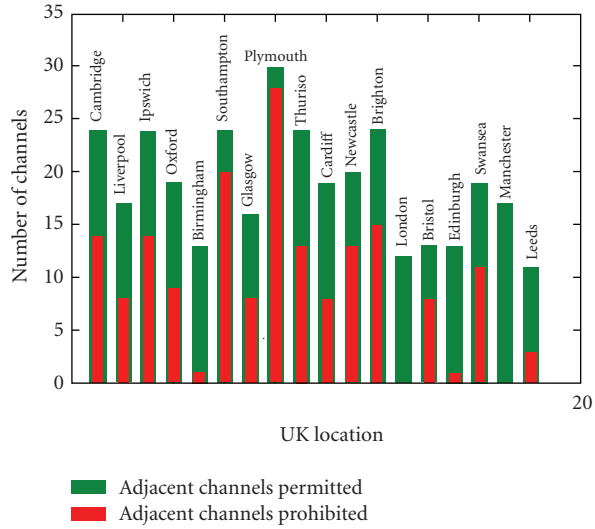


Figure 4: Available TVWS capacity for low-power cognitive access in 18 UK locations as obtained from coverage modelling. Results are shown both without (green bars) and with (red bars) considering adjacent channel interference constraint.

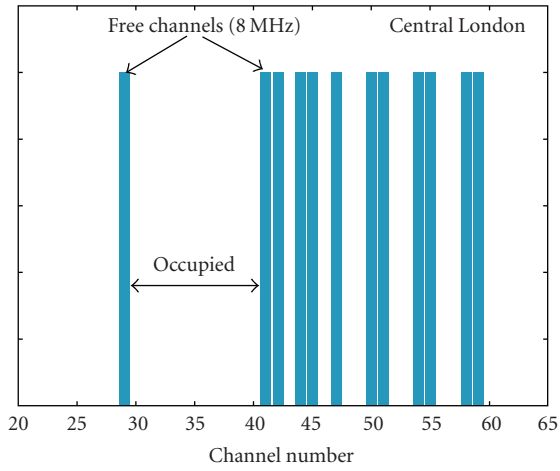


Figure 5: TVWS channels available for cognitive access in Central London.

study performed in [14] shows that in the US the main channels of relevance are the lower UHF channels where ~15 (90 MHz) channels per location/per person are available for low power cognitive access. However this number drops significantly (to ~5) when adjacent channels also have to be protected.

In addition to estimating total available TVWS, it is of importance to investigate channel composition of this spectrum. In Figure 5 we show, as an example, channel composition of TVWS in Central London. In this figure vacant channels are shown as blue bars while occupied channels are left black. As can be seen from the figure, the available TVWS channels can be highly non-contiguous. This feature may greatly restrict access to TVWS by most

current wireless technologies, as modulation schemes implemented in these technologies often require a contiguous portion of the spectrum. In the case of London although a total of 96 MHz spectrum is in principle available, only 16 MHz can be utilised for contiguous frequency access.

We note that the above results were obtained using a combination of highly realistic DTV coverage maps with a simplified pathloss propagation model, which is adequate when considering low-power cognitive devices. Although our results are in good agreement with other independently obtained modelling results, for example, by Ofcom, we are currently not aware of any actual measurements of TVWS avilability in the UK that could be used to directly validate our modelling results. Furthermore, in order to improve the accuracy of our TVWS estimates for high-power cognitive devices, such as those required in rural broadband applications, we are incorporating standard UHF propagation models such as the ITU-R P.1546-3 (Longley-Rice) model [41] into our calculations.

4.2. Use Cases. In addition to considerable capacity it offers, which is evident from the discussion in the previous section, an important reason why TVWS spectrum has attracted much interest is an exceptionally attractive combination of bandwidth and coverage. Signals in the VHF/UHF TV bands travel much further than both the WiFi and 3G signals and penetrate buildings more readily. This in turn means that these bands can be used for a very wide range of potential new services. In addition to broadband wireless access to undeserved areas, other technologically important applications of TVWS spectrum include the following:

(i) wireless distribution networks for future digital homes and smart energy grids,

(ii) licensed-exempt mobile broadband,

(iii) last mile wireless broadband in urban environments,

(iv) cognitive femtocells/cellular communications in TVWS.

In the following paragraphs we will focus our attention on two future use cases of cognitive access to TVWS, which have recently attracted much attention from both industry and research community.

4.2.1. Future Wireless Home Networks. Fuelled by the quick progress of wireless technologies, broadband adoption, and without the burden of spectrum licences, home wireless networking has become in the last few years a pervasive technology. Between 2004 and 2006, home network adoption boomed across Europe, with growth rates surpassing Asia and North America. France and the UK both trebled the number of households with a home network, putting them slightly ahead of the US Italy and Germany still lagged behind but posted notable growth nonetheless. More than 54% of European households have a computer and a total of 34% are using WiFi routers. The future wireless home will consist not only of PC, laptops and PDAS wirelessly connected to the Internet but also media servers (High Definition TV, video and audio), access points, computer electronics like wireless cameras and game consoles. Due to sharp decrease in wireless solutions prices and minimum restrictions for access to ISM bands, other domestic applications as well as gas, electricity and water meters will in the future come equipped with radio receivers allowing control, monitoring and easy configuration (the so-called smart grids).

Most of these devices and services support wireless connectivity using one or a number of short-range wireless technologies, such as WiFi (IEEE 802.11), Zigbee (IEEE 802.15.4), and so forth, all operating without the need for a licence in the already congested ISM bands. Home networks of the future operating exclusively in these bands are expected to suffer severe capacity limitations resulting from interference caused by the high device density and limited spectrum availability in the ISM bands. Furthermore, the aggregate interference resulting from these devices is bound to create a high interference burden on the WiFi-based provision of broadband wireless access in homes.

Additional capacity offered by secondary access to TVWS has the potential to solve this capacity limitation problem thereby contributing to increasing takeup of wireless home networking and services, and spurring future technological innovation and revenue generation. In particular, some of the most bandwidth-intensive home networks applications (such as multimedia streaming) can be offloaded to TVWS bands hence freeing up the ISM bands for other consumer applications. Our recent system-wide simulation studies show that [13] due to lower operation frequencies home access points operating in the TVWS UHF frequencies can achieve throughput levels that are either higher or comparable to WiFi access points while using significantly lower transmit power levels (two order of magnitudes in mW) [13]. An additional benefit here is that a significant saving in energy consumption can be achieved in home networking scenarios by switching from the ISM bands to TVWS bands [13]. Furthermore, according to several recent studies [12, 42] the number of available channels for indoor white space transmissions appear to be very significant.

Protection of incumbents in such home networking scenarios can be achieved using a master-slave architecture where functionalities of spectrum detection and/or geolocation and database access and spectrum assignment are all integrated into the home access point [43]. The access point monitors the availability of spectrum in the ISM and TV bands and instruct customer devices which spectrum to use based on their bandwidth and QoS requirements.

4.2.2. Cognitive/TVWS Femtocells. A femtocell is a small base station of 3G/LTE, or WiMAX technology, controlled by the mobile operator and placed inside the home /small office of the customer. Femtocells are useful when a user experiences bad indoor coverage or its application is too capacity-demanding for indoor conditions. The user may be already inside or going inside a building. Femtocells help maintaining a mobile broadband session or to allow it where it previously was not possible. Current generation femtocells use the same frequencies as mobile networks, hence creating

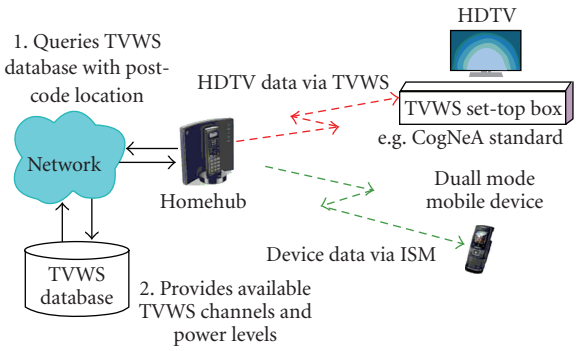


Figure 6: Architecture for HDTV distribution using TV White Space spectrum.

a potential source of interference that can be difficult to control since the user femtocell is not controlled by the operator.

Femtocells operating in TVWS would be an alternative to femtocells proprietary technologies that are appearing on the market for dedicated 3G/LTE networks [44]. The main advantage of CR based femtocells compared to traditional femtocells will be reduced or better controlled interference into the operators' network. Another case of great interest to operators is to use CR to backhauling of femtocells (either traditional or CR-based themselves). This allows a mobile operator gain control into the home of the user, should it be outside the DSL coverage or whether the user has another operator delivering broadband to his/her home network.

5. Conclusions

In this paper we surveyed the state-of-the-art in cognitive radio access to TV White Spaces. We showed that a regulatory framework for secondary utilisation of TVWS spectrum is well underway both in the US and UK and important steps in this direction are also being taken within the EU and worldwide. Using result from recent quantitative studies of the TVWS availability in the UK and US we illustrated that cognitive access to these bands provides a very significant spectrum opportunity for a range of indoor and outdoor applications and services. In addition to rural broadband, which is the main focus of the IEEE 802.22 standard, these include wireless home networks, mobile broadband, and TVWS femtocells.

However, effective exploitation of this spectrum for such commercial services requires addressing an array of important technology challenges. One of these, high-precision spectrum sensing, has been the subject of numerous research papers while others, including multiple secondary access, aggregate interference control, and agile modulation techniques, have not yet received the attention we believe they deserve. Furthermore, quantitative techno-economical studies of the commercial feasibility and cost versus benefit associated with use cases of cognitive radio crucial in influencing the takeup of the technology by wireless network and service providers but are currently very limited. Our own research is currently focusing on some of the abovementioned technology and business challenges of cognitive access to TVWS.

Acknowledgments

The author wishes to acknowledge his colleagues at BT, S. Kawade, M. Fitch, X. Gu, K. Briggs, R. MacKenzie, P. Bruce, A. Sago, C. Cheeseman, M. Pine, J. Zhang, N. Edwards, and D. Bryant, for stimulating discussions. He thanks Jonas Kronander from Ericsson AB for reading the manuscript. The author would like to thank the European Union for providing partial funding of this work through the EU FP7 project INFSO-ICT-248303 QUASAR.

References

[1] A. Wyglynski, M. Nekovee, and T. Hou, Eds., *Cognitive Radio Communication and Networks: Principle and Practice*, Academic Press, London, UK, 2010.

[2] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[3] M. Nekovee, "Dynamic spectrum access—concepts and future architectures," *BT Technology Journal*, vol. 24, no. 2, pp. 111–116, 2006.

[4] Federal Communications Commission, "Spectrum policy task force," Report of the Spectrum Efficiency Working Group, November 2004, http://www.fcc.gov/sptf/reports.html/.

[5] Federal Communications Commission (FCC), "Second report and order and memorandum. Opinion and order," Tech. Rep. 08-260, November 2008.

[6] Federal Communications Commission (FCC), "Additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band," ET Docket 02-380, December 2002.

[7] Federal Communications Commission (FCC), "Unlicensed operation in the TV broadcast bands," ET Docket 04-186, May 2004.

[8] Ofcom, "Digital Dividend Review, A statement on our approach towards awarding the digital dividend," December 2007.

[9] "Ofcom Consultation: Digital Dividend," February 2009, http://www.ofcom.org.uk/consult/condocs/cognitive/summary/.

[10] Ofcom, "Statement on Cognitive Access to Interleaved Spectrum," July 2009.

[11] J. Bradford, T. Cook, D. Ramsbottom, and S. Jones, "Optimising usage of spectrum below 15 GHz used for defence in the UK," in *Proceedings of the IET Seminar on Software Defined Radio and Cognitive Radio*, London, UK, September 2008.

[12] M. Nekovee, "Quantifying the availability of TV white spaces for cognitive radio operation in the UK," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '09)*, Dresden, Germany, June 2009.

[13] S. Kawade and M. Nekovee, "Wireless options for high data-rate indoor users: cognitive access to TV white spaces," in *Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10)*, Singapore, April 2010.

[14] M. Mishra and A. Sahai, "How much white space is there?" Tech. Rep., Electrical Engineering and Computer Sciences, University of California at Berkeley, January 2009.

[15] T. X. Brown and D. C. Sicker, "Can cognitive radio support broadband wireless access?" in *Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '07)*, pp. 123–132, April 2007.

[16] "CogNeA," http://www.cognea.com/.

[17] K. Challapali, "Philips Research North America (private communication)," FAQ on CogNeA Alliance, K. Kimyacioglu, Philips.

[18] Ecma 392, Ecma International Standard.

[19] J. Wang, M. Sun Song, S. Santhiveeran, et al., "First cognitive PHY/MAC standard for personal/portable devices in TV White Spaces".

[20] "IEEE 802.22," http://www.ieee802.org/22/.

[21] C. Cordeiro, D. Cavalcanti, and S. ShankarA. Wyglynski, M. Nekovee, and T. Hou, "Cognitive radio for broadband wireless access in TV bands: the IEEE 802.22 standards," in *Cognitive Radio Communication and Networking: Principle and Practice*, Academic Press, Boston, Mass, USA, 2009.

[22] A. Sahai and D. Cabric, "Spectrum sensing: fundamental limits and practical challenges," in *Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, Baltimore, Md, USA, November 2005.

[23] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[24] H.-S. Chen, W. Gao, and D. G. Daut, "Spectrum sensing for wireless microphone signals," in *Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops*, San Francisco, Calif, USA, June 2008.

[25] A. Mossa and A. Jeoti, "Cyclostationary-based spectrum sensing for analog TV and wireless microphone signals," in *Proceedings of the 1st International Conference on Computational Intelligence, Communication Systems and Networks*, pp. 380–385, 2008.

[26] P. D. Sutton, K. E. Nolan, and L. E. Doyle, "Cyclostationary signatures in practical cognitive radio applications," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 13–24, 2008.

[27] L. P. Goh, Z. Lei, and F. Chini, "DVB detector for cognitive radio," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, pp. 6460–6465, Glasgow, UK, June 2007.

[28] G. Noh, J. Lee, H. Wang, S. You, and D. Hong, "A new spectrum sensing scheme using cyclic prefix for OFDM-based cognitive radio systems," in *Proceedings of the IEEE Vehicular Technology Conference (VTC '08)*, pp. 1891–1895, May 2008.

[29] "BT Response to Ofcom Consultation on Digital Dividend," Cognitive Access.

[30] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 131–136, Baltimore, Md, USA, November 2005.

[31] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: two user networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2204–2212, 2007.

[32] Y. Selén and J. Kronander, "Cooperative detection of Programme making special event devices in realistic fading environements," in *Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10)*, Singapore, April 2010.

[33] R. Rajbanshi, A. M. Wyglinski, and G. J. Minden, "An efficient implementation of NC-OFDM transceivers for cognitive radios," in *Proceedings of the 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '06)*, June 2006.

[34] R. Rajbanshi, *OFDMbased cognitive radio for DSA networks*, Ph.D. thesis, University of Kansas, Lawrence, Kan, USA, 2007.

[35] W. Rhee, J. C. Chuang, and L. J. Cimini, "Performance comparison of OFDM and multitone with polyphase filter bank for wireless communications," in *Proceedings of the 48th IEEE Vehicular Technology Conference (VTC '98)*, pp. 768–772, May 1998.

[36] M. Jankiraman, *Space-Time Codes and MIMO Systems*, Artech House, Boston, Mass, USA, 2004.

[37] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *IEEE Communications Magazine*, vol. 44, no. 6, pp. 44–49, 2006.

[38] N. S. Shankar and C. Cordeiro, "Analysis of aggregated interference at DTV receivers in TV bands," in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '08)*, Singapore, May 2008.

[39] "UK FREE.TV," http://www.ukfree.tv/.

[40] K. Briggs, "BT innovate and design," Tech. Rep., July 2009.

[41] P. L. Rice, A. G. Longley, K. A. Norton, and A. P. Barsis, "Transmission loss predictors for tropospheric communication circuits," Tech. Note 101, U S Government Printing Office, Washington, DC, USA, 1965.

[42] E. Obregon and J. Zander, "Short range white space utilization in broadcast systems for indoor environments," in *Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10)*, Singapore, April 2010.

[43] M. Nekovee, "Cognitive access to TV White Spaces: spectrum opportunities, commercial applications and remaining technology challanges," in *Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10)*, Singapore, April 2010.

[44] D. López-Pérez, A. Valcarce, G. De La Roche, and J. Zhang, "OFDMA femtocells: a roadmap on interference avoidance," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 41–48, 2009.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 905321, 9 pages
doi:10.1155/2010/905321

Research Article

Backpropagation-Based Cooperative Localization of Primary User for Avoiding Hidden-Node Problem in Cognitive Networks

Lin Liu, Zhengyi Li, and Chi Zhou

Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 South Dearborn Street, Chicago, IL 60616, USA

Correspondence should be addressed to Chi Zhou, zhouc@ece.iit.edu

Received 29 November 2009; Revised 24 March 2010; Accepted 28 April 2010

Academic Editor: Hsiao Hwa Chen

Cognitive radio (CR) is a technology to implement opportunistic spectrum sharing to improve the spectrum utilization. However, there exists a hidden-node problem, which can be a big challenge to solve especially when the primary receiver is passive listening. We aim to provide a solution to the hidden-node problem for passive-listening receiver based on cooperation of multiple CRs. Specifically, we consider a cooperative GPS-enabled cognitive network. Once the existence of PU is detected, a localization algorithm will be employed to first estimate the path loss model for the environment based on backpropagation method and then to locate the position of PU. Finally, a disable region is identified taking into account the communication range of both the PU and the CR. The CRs within the disabled region are prohibited to transmit in order to avoid interfering with the primary receiver. Both analysis and simulation results are provided.

1. Introduction

As more devices go wireless, spectrum becomes more and more crowded. Study of spectrum utilization, however, reveals that not all the spectrum is in use for all the time. Enforcement Bureau of Federal Communications Commission (FCC) measures the spectrum usage in Atlanta, Chicago, and so forth, and the study shows that only 5%–10% of the spectrum is used (up to 100 GHz) on the average. DARPA study reveals that only 2% of the allocated spectrum is used at any given time. Therefore, there is a potential to make efficient use of the unused spectrum without interfering with primary users (PUs) so that the spectrum utilization can be improved and more users can be supported. Cognitive radio (CR) is a technology to implement opportunistic spectrum sharing to improve the spectrum utilization [1–3]. CR can be applied in civilian applications, law enforcement, as well as military applications.

For CR, spectrum sensing is the first step but very crucial to the success. Only when the electromagnetic environment is thoroughly understood, it can be decided over which frequency to transmit and how to transmit. As the cognitive radio is seen as the secondary user to share the licensed band with the PU, they must avoid or control the interference to potential PU. However, as a radio device, a single CR may suffer severe shadowing or multipath fading with respect to primary transmitter so that it cannot detect the existence of primary transmitter even in its vicinities. In addition, there exists a hidden-node problem, in which a CR may be too far from the transmitter to detect the existence of the PU, but close to the primary receiver to interfere with the reception if transmit. Cooperative sensing provides a solution to the challenges mentioned above [4, 5]. In cooperative sensing, multiple cognitive radios cooperate to reach an optimal global decision by exchanging and combining individual local sensing results. Allowing multiple CRs to cooperate, cooperative sensing can increase the detection probability, reduce the detection time, and achieve the diversity gain [6–11].

In this paper, we aim to provide a solution to the hidden-node problem for passive-listening receiver based on cooperation of multiple CRs. Solutions to hidden-node problem have been provided, such as RTS/CTS handshake for WLAN [12], BTMA (Busy Tone Multiple Access) for a centralized system [9], and DBTMA (Dual Busy Tone Multiple Access) for ad hoc networks [10]. However, most of

the solutions require active participation of primary receiver and fail when primary receiver is just passive listening. When passive listening, primary receiver does not acknowledge or respond. In our approach, GPS-enabled CRs cooperate to first estimate the environment based on back propagation method and then locate the position of PU. Based on the PU location derived, we identify the disable region, whose radius is the sum of PU communication range and a CR communication range. Within the disable region, all the CRs are prohibited to transmit in order to avoid interfering with the PU. Since the disable region takes into account the CR communication range, our proposed method deals with the worst case of hidden-node problem. In addition, it does not require any involvement from primary receiver, suitable for passive-listening scenario.

The remainder of the paper is organized as follow. In Section 2, we present our system model. In Section 3, we propose the localization algorithm to first estimate the path loss model and further determine the exact location of PU. Simulation results are provided in Section 4, and finally the paper is concluded by Section 5.

2. System Model

We consider a cognitive network, which consists of N CRs that are GPS enabled and cooperative, shown in Figure 1. A simple energy detector is used by individual CR to detect the existence of PU by measuring the strength of the received power. A central control office collects the sensing results from individual CR to make a global detection decision. If the existence of PU is determined, the central control office runs a localization algorithm to first estimate the path loss model for the environment and then to locate the position of PU.

Once the position of PU is located, a disable region for CRs can be identified. Let R_p denote the communication range for PU and R_c denote the communication range for CR, and then the disable region is the circle area with the origin at the PU location and the radius $R_d = R_p + R_c$. The actual CR communication range depends on the CR receiver and the CR transmits power jointly, which are known to the system. Therefore, it can be easily calculated given the path loss model. Finally, the central control office sends the control signal to individual CR to either disable or enable the CR depending on whether it locates within or outside of the disable region. The CRs within the disabled region are prohibited to transmit in order to avoid interfering with the primary receiver, while CRs outside of the region can transmit. Since the disable region takes into account the CR communication range, the hidden-node problem can be solved.

Widely adopted AODV (Ad Hoc On-Demand Distance Vector) routing protocol is used over a default clear channel to exchange the information between the CRs and the central control office. AODV protocol has small delay close to 8 ms in general and 20 k to 30 k routing overhead [13]. The default channel may be selected among several predetermined channels. A clear channel can be easily identified from the spectrum sensing stage.

In this paper, we use the log-normal shadowing path loss model:

$$P_r(d) = P_0(d_0) + 10 \cdot n \cdot \log\left(\frac{d_0}{d}\right) + X(0, \delta), \tag{1}$$

where $P_r(d)$ is the received power at distance d, $P_0(d_0)$ is the received power at the reference distance d_0, n is the path loss exponent, and $X(0, \delta)$ is normal shadowing random variable with zero mean and δ variance.

3. Localization Algorithm

In this section, we propose a localization algorithm, which will be employed to estimate the path loss model and further locate the position of PU once the presence of PU is detected. Since all CRs are GPS-enabled, the central control office can obtain the location information from those CRs. If we know the relative distances between PU and CRs, we can calculate the position of PU easily. As shown in (1), the distance between PU and a CR can be estimated by the received power for that CR and the path loss exponent n. Therefore, we need to estimate the path loss exponent first and then locate the PU.

3.1. Localization Algorithm with Identical Path Loss Coefficients. We first consider the simple scenario in which all CRs locate in the same electromagnetic environment, that is, they have the same path loss exponent, expect for those CRs located close to PU, whose path loss should be modeled as free space path loss model. In (1), the reference point d_0 is always chosen in the close vicinity of PU; therefore, $P_0(d_0)$ can be calculated from free-space equation

$$P_0(d_0) = \frac{P_t \cdot G_t \cdot G_r \cdot {\lambda_f}^2}{(4 \cdot \pi)^2 \cdot {d_0}^2}, \tag{2}$$

where P_t is the transmit power from the transmitter, G_t and G_r are the antenna gains for transmitter and receiver, respectively, and λ_f is the wavelength of the carrier frequency.

Our localization algorithm is specified in details as follows (shown in Figure 2).

Step 1. Sort all CRs according to the descending order of the received power and number the CRs from 0 to $N - 1$.

Step 2. If the received power of CR0 is above certain threshold, we pick CR0 as the reference, calculate the reference distance, and then continue the process. Otherwise, we will repeat Steps 1 and 2 till we find CR0 with received power exceeding the threshold.

That is, we treat the location of CR0 (x_0, y_0) as the reference point and the received power of CR0 as $P_0(d_0)$. Then the reference distance d_0 can be calculated according to (2) as

$$d_0 = \sqrt{\frac{P_t G_t \cdot G_r \cdot {\lambda_f}^2}{P_0(d_0)(4 \cdot \pi)^2}}. \tag{3}$$

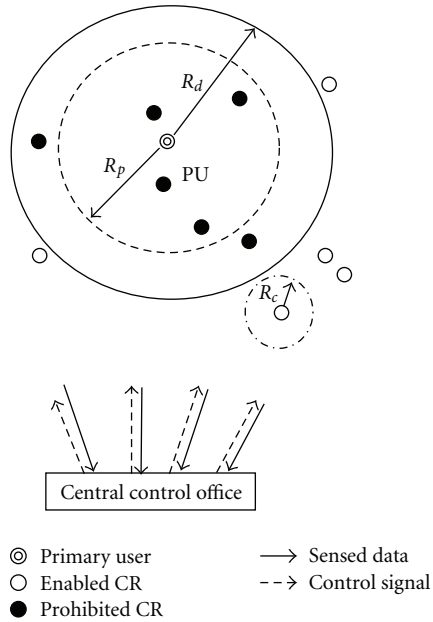


Figure 1: System model.

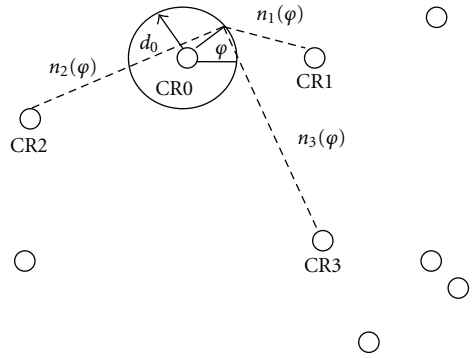


Figure 2: Illustration of Localization Algorithm.

From spectrum sensing stage, the carrier frequency of the PU can be identified. The transmit power is usually regulated to certain level. In the United States, FCC regulates the use of antennas not to exceed certain power limitations for wireless devices using the parameter EIRP (equivalent isotropically radiated power), which represents the effective transmit power of the radio in total, including transmit antenna gains. When using omnidirectional antennas (less than 6 dB), the FCC rules require EIRP to be 1 watt or less. In this paper, we use $P_t G_t = 1$ watt.

The choice of the power threshold depends on how much accuracy or estimation error the system can tolerate. One example is illustrated in Figure 4.

Step 3. Draw a circle with the CR0 location as the center and d_0 as the radius. If CR0 does follow the free-space propagation, the PU must locate on that circle.

Step 4. Pick the next three CRs, CR1, CR2, and CR3, in the list. Move the PU location along the circle from 0-degree angle to 360-degree angle and calculate the corresponding path loss exponent n for CR1, CR2, and CR3, respectively.

For example, we consider CR1. When the PU locates on the circle with angle φ, the PU location is $(x_0 + d_0 \cos\varphi, y_0 + d_0 \sin\varphi)$, and the distance between the point and CR1 is

$$d_1 = \sqrt{(x_0 + d_0 \cos\varphi - x_1)^2 + (y_0 + d_0 \sin\varphi - y_1)^2}. \quad (4)$$

According to (1), when ignoring the shadowing, the path loss exponent is

$$n_1(\varphi) = \frac{P_0(d_0) - P_1}{10 \log(d_1/d_0)}. \quad (5)$$

Step 5. Calculate the relative difference between the path loss exponents among those three CRs for any angle φ:

$$\begin{aligned} E12 &= |\, n_1(\varphi) - n_2(\varphi) \,|, \\ E13 &= |\, n_1(\varphi) - n_3(\varphi) \,|, \\ E32 &= |\, n_3(\varphi) - n_2(\varphi) \,|. \end{aligned} \quad (6)$$

We pick the angle that gives the minimum difference summation:

$$\overline{\varphi} = \arg\min(E12 + E13 + E32). \quad (7)$$

Then the path loss exponent $\overline{n}$ can be estimated as

$$\overline{n} = \frac{n_1(\overline{\varphi}) + n_2(\overline{\varphi}) + n_3(\overline{\varphi})}{3}. \quad (8)$$

The position of PU can be estimated as

$$\begin{aligned} \hat{x} &= x_0 + d_0 \cos\overline{\varphi} \\ \hat{y} &= y_0 + d_0 \sin\overline{\varphi}. \end{aligned} \quad (9)$$

Let $P_{\min}$ be the minimum acceptable received signal power for the main receiver from PU. Ignoring the random shadowing, we can calculate the communication range of PU as

$$R_p = \frac{10^{(P_0(d_0) - P_{\min})/10\overline{n}}}{d_0}. \quad (10)$$

And then the radius for disable region for CRs is

$$R_d = R_p + R_c. \quad (11)$$

Without shadowing, our algorithm can accurately estimate the path loss exponent and locate the exact location of PU. When there is shadowing, the shadowing effect can be reduced if we take multiple samples of the received power to average out the randomness.

CR0, though having the strongest received power, may not follow the free-space path loss model. In this case, we need to evaluate the proposed localization algorithm in terms of error, which is defined as

$$\text{Error} = \frac{d_e}{R_p} = \frac{\sqrt{(x_{\text{PU}} - \hat{x})^2 - (y_{\text{PU}} - \hat{y})^2}}{R_p}, \tag{12}$$

where d_e is the separation distance between the actual PU and the estimated PU, $(x_{\text{PU}}, y_{\text{PU}})$ is the actual PU location, and $(\hat{x}, \hat{y})$ is the estimated PU location. Some simulation results will be given in Section 4.

3.2. Localization Algorithm with Variant Path Loss Coefficients. It is likely that the CRs locate in different environment; thus, we propose a backpropagation-based algorithm to locate the PU under a variant environment. Back propagation has been used in various areas, such as the artificial neural network and the MIMO process, but no work has been found in communication networks to the best of our knowledge.

As the algorithm's name implies, the errors (and therefore the learning) propagate backwards from the output nodes to the inner nodes in a network. So technically speaking, back propagation is used to calculate the gradient of the error of the network with respect to the network's modifiable weights. This gradient is always then used in a simple stochastic gradient descent algorithm to find weights that minimize the error. Often the term "back propagation", used in a more general sense, refers to the entire procedure encompassing both the calculation of the gradient and its use in stochastic gradient descent. Back propagation usually allows quick convergence on satisfactory local minima for error in some appropriate network settings.

In this paper, we can consider our model as a 2-input/1-output system. All we have to do is to estimate all the weights and hidden nodes in the system to approximate the true results. According to (1), there are two unknown variables in the equation: d and n. In the variant environment, path loss coefficients can be varied. But we can first estimate the rough position of PU using AOA estimation method. This result will not provide us an acceptable result. However, it can eliminate one variable in (1). In this way, we can employ the back propagation algorithm to find out the path loss coefficients of all the cognitive radios. Using the results of back propagation algorithm, we can further estimate the exact location of the PU.

For back propagation algorithm, we can treat (1) in the form as below:

$$P_r = f(n, d). \tag{13}$$

The received power calculated from (13) varies from the true value of the received power depending on the values of the path loss coefficients and the estimated distances. As we talk above, we can find a way to first estimate the location of the PU with a larger error. In order to get the closest result to the true value, we introduce (14) to calculate the square error. Finding a good estimation of the path loss coefficients will now become finding the minimum value of function $J(n)$:

$$J(n) = \left(f(n, d) - P_{\text{real}}\right)^2, \tag{14}$$

where P_{real} is the actual measured received power.

We start finding the proper path loss coefficients with a set of random initial values. Then we process those variables with the back propagation algorithm. The algorithm is described as shown in the following steps

Step 1 *and Step* 2. are the same as those in localization algorithm with the identical path loss.

Step 3. CR0 estimates the angle of arrival (AOA), θ, of the received signal. The rough position of the PU can be calculated as

$$\begin{aligned} \hat{y}_{\text{Initial}} &= y_0 + d_0 \sin\theta, \\ \hat{x}_{\text{Initial}} &= x_0 + d_0 \cos\theta. \end{aligned} \tag{15}$$

Step 4. Using the back propagation algorithm [14] to obtain the path loss coefficients for every cognitive radio. The detail process can be described as follows

(1) Pick initial values of the path loss coefficient for each CR using pseudorandom process that randomly selects n_i within the range from 2 to 8 for the ith CR.

(2) Compute the receive power of each CR with the coefficients according to (13).

(3) Compute the objective function

$$J(n_1, \ldots, n_M) = \frac{1}{M} \sum_{i=1}^{M} \left(f(n_i, d_i) - P_{i,\text{real}}\right)^2, \tag{16}$$

where M is the number of the CRs that we use in the algorithm, $f(n_i, d_i)$ is the estimated received power calculated using the selected n_i for the ith CR, and $P_{i,\text{real}}$ is the actual measured received power obtained by the antenna of the ith CR, n_1 to n_M is the path loss exponent that we need to obtain in this step.

(4) Take the second-order derivative of $J(n_1, n_2, \ldots, n_M)$ with respect to each path loss coefficient n_i and select the smallest value as the updating step η:

$$\eta = \min\left(\frac{\partial J^2(n_1, \ldots, n_M)}{\partial n_1^2}, \frac{\partial J^2(n_1, \ldots, n_M)}{\partial n_2^2}, \ldots, \frac{\partial J^2(n_1, \ldots, n_M)}{\partial n_M^2}\right). \tag{17}$$

(5) Update the parameters n_i using the following equation:

$$n_i^{L+1} = n_i^L + \eta \frac{\partial J(n_1, \ldots, n_M)}{\partial n_i}, \tag{18}$$

where L denotes the iteration number.

(6) Repeat procedures (2) to (5) until we have a decent estimation. Usually we have an acceptable answer after 70 iterations.

Step 5. Compute distance between the PU and every CR using the equation given below:

$$d_i = \frac{d_0}{10^{(P_{i,\text{real}} - P_o(d_o))/10\hat{n}_i}}. \tag{19}$$

Step 6. Determine the area that the PU might locate, which is specified by the intersections of each circle centered at each CR with the estimated d_i from Step 5 as the radius.

Step 7. Compute a 3D probability contour and pick the point giving us the largest probability as the PU location. The calculation of the probability for point c on the intersection plane is given as (20):

$$\text{Prob} = \prod \text{pdf}\left\{N\left(P_{i,\text{real}} - Pd_0 - 10\hat{n}_i \cdot \log\left(\frac{d_0}{d_{c,i}}\right), \delta\right)\right\}, \tag{20}$$

where $d_{c,i}$ is the distance between point c and the ith CR.

4. Simulation Results

4.1. Localization Simulation with Identical Environment. We consider a cooperative GPS-enabled cognitive network with the following parameter setup: PU location: (22, 23), CR0 location: (22, 22), CR1 location: (6, 5), CR2 location: (40, 6), CR3 location: (7, 41), $P_tG_t = 1$ watt, $G_r = 1$, and carrier frequency is 300 MHz. In all simulations, we choose −4.5 dbm as the power threshold (illustrated in Figure 4).

We further assume that the path loss for CR0 follows free-space equation, and CR1, CR2, and CR3 have the same path loss exponent, $n = 4$. We consider two scenarios, without shadowing and with shadowing.

4.1.1. Without Shadowing. In this case, the path loss model is simplified to log-distance model. This is a deterministic model without randomness. We plot the corresponding path loss exponent estimate for CR1, CR2, and CR3, respectively, as we move the potential PU point along the circle from 0 degree to 360 degrees.

From Figure 3, we can see that the path loss exponent varies as the point moves, as the distance between the potential PU and CR changes. In addition, the three curves cross at a single point with the angle $\varphi = 90$ degrees and the corresponding path loss exponent $n = 4$. According to (6) to (8), the crossing point has the minimum summed differences; therefore, we have $\overline{\varphi} = 90$ degrees and path loss exponent $\overline{n} = 4$. That is, the estimations perfectly match the exact values.

We then examine how to choose the power threshold in step 2 to determine whether a CR can be considered as the reference. We assume that the reference distance is $d_0 = 1$ and the actual path loss for the CR is 4. In general, the path loss exponent varies between 2 and 4. Therefore, we consider the worst case. We vary the distance between the PU and the CR and calculate the corresponding error using (12). The results are illustrated in Figure 4. It is shown that the error increases with the distance. If we choose the 5% error, the distance is 2 and the corresponding received power is −4.5 dbm, which is the chosen power threshold.

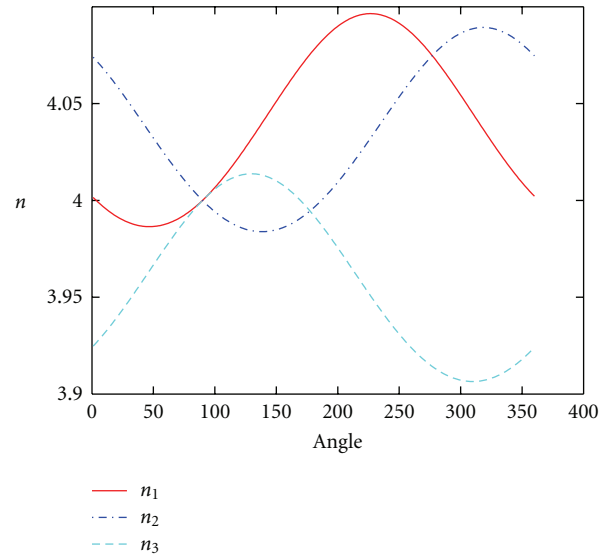


Figure 3: Path loss exponent versus angle without shadowing.

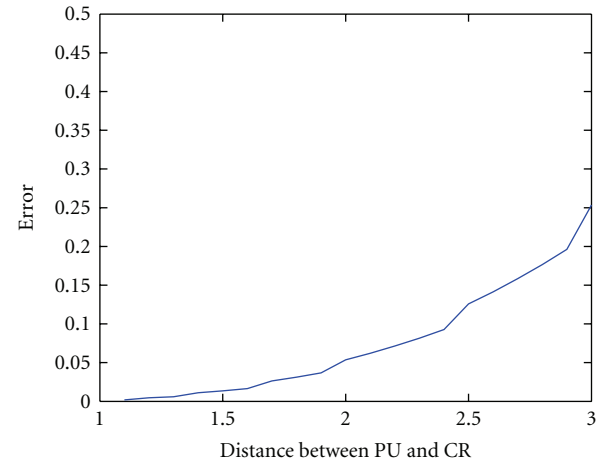


Figure 4: Error versus the distance between the PU and CR0.

4.1.2. With Shadowing. Then we add random shadowing into the path loss model. We first set the shadowing random variable with zero mean and 8 dB variance. To average out the randomness, we take 600 samples and average the received power, P_1, P_2, and P_3. We redraw the path loss exponent curves as we move the potential PU point along the circle from 0 degree to 360 degrees, shown in Figure 5. The three curves do not intersect at a single point in contrast to the no-shadowing case. Again, we use (6) to (8) to estimate the angle $\overline{\varphi}$ and path loss exponent $\overline{n}$. It is shown that at 99 degrees the three exponents have the minimum summed difference, that is, $\overline{\varphi} = 99$ degrees and $\overline{n} = 4.05$. Then the estimated location for PU is (21.8491, 22.9529). As we can see, the estimated position is very close to the exact location of PU.

We would like to see how the increase of the sample number would affect the performance. Therefore, we increase the number of samples, and check the estimated location of PU. Figure 6 shows how the estimated x-coordinate varies

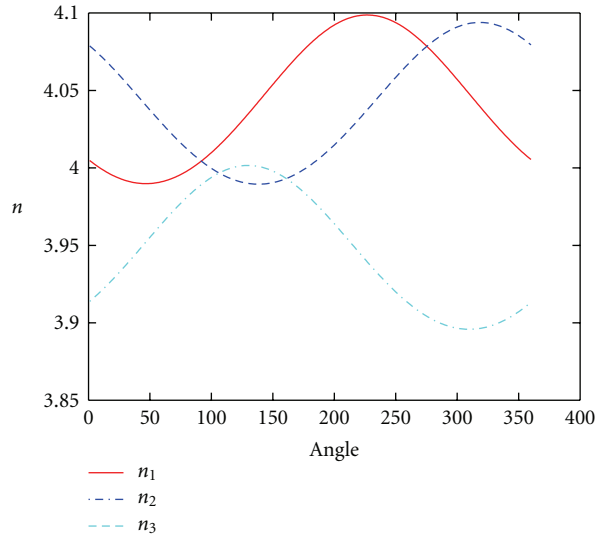


Figure 5: Path loss exponent versus angle using 600 samples with shadowing.

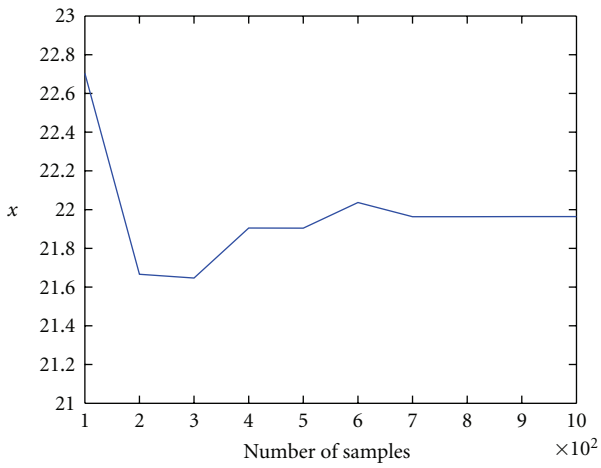


Figure 6: x coordinate of PU location versus sample number.

Table 1

Sample number	100	200	300	400	500
Angle	65	134	77	63	71
Sample number	600	700	800	900	1000
Angle	80	92	92	92	92

with the number of samples. It is shown that the estimated value approaches to the exact value as the number of samples increases. In this specific example, we observe that the curve reaches the maximum error at 200 samples; however, the difference is still less than 1, which is less than 5% as an error. In addition, the curve approaches to the exact 22 as the number of samples increases. The same property can be found for y coordinate.

We also estimate the angle given the various number of samples, as shown in Table 1.

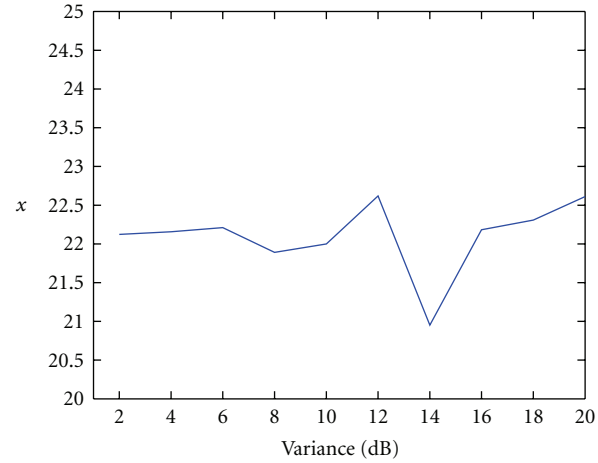


Figure 7: x-coordinate of PU location versus variance of shadowing.

As we can observe here, the estimated angle also approaches to the exact value 90 of degrees as the number of samples increases. When more than 800 samples are used, we can get the estimated angle really close to the exact value of 90 degree with only 1 or 2 degrees difference.

In addition, we want to study the effect of the shadowing variance casting on the accuracy of the localization. In this simulation, we use 1000 samples for each outcome under different variance.

Figure 7 shows the estimated x-coordinate for different shadowing variance. We can see that the difference between the exact value and estimated value tends to become larger as we increase the variance of shadowing. If we choose the variance less than 10 dB, we can estimate the position of PU within small error using the proposed localization algorithm. Similar properties are found for y-coordinate and the estimated angle.

4.2. Localization Simulation with Variant Environment. In this section, we focus on the performance of the localization algorithm with variant environment. We assume that PU location (23, 23) and path losses for CR1, CR2, and CR3 are 5, 4, and 6, respectively, while keeping other parameters unchanged as A with the shadowing random variable with zero mean and 8 dB variance.

The back propagation algorithm is applied to estimate the path loss coefficients. Figure 8 shows the estimated coefficients versus the number of iterations. It is shown that the back propagation algorithm takes approximately 100 iterations to reach the steady final result. Note that the number of iteration needed varies with the distance between the PU and the reference point.

Once the path loss coefficients are estimated, we process Step 6 to determine in which area the PU may locate. We calculate the estimated distance between the PU and each cognitive radio and then draw a circle which indicates the probable area of the PU. Then we take all the intersections and overlaps into consideration and identify the area for PU.

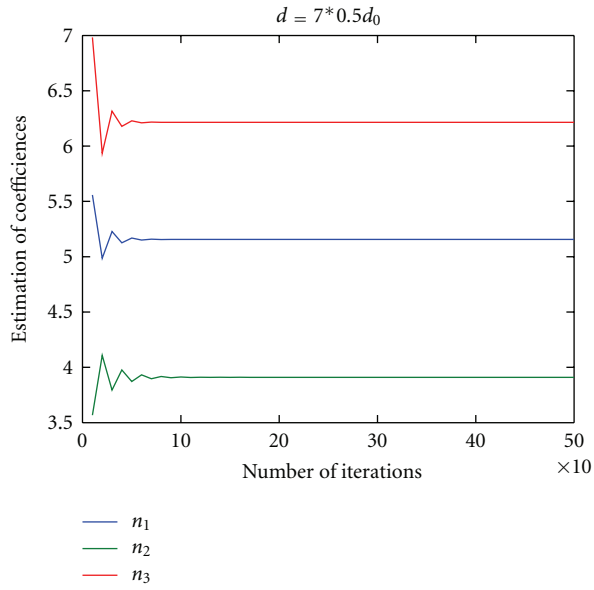


Figure 8: Estimation of path loss coefficients versus iterations number.

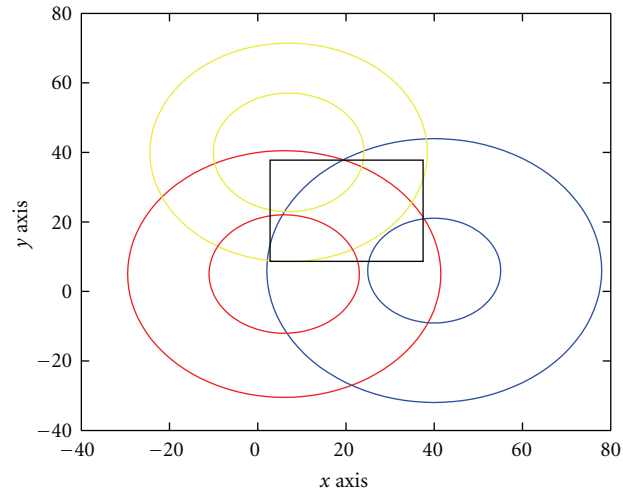


Figure 9: Probable area for the location of PU.

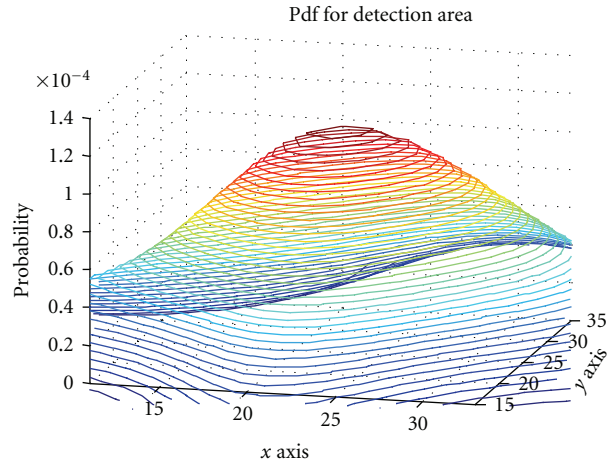


Figure 10: 3-D probability contour of the detection area.

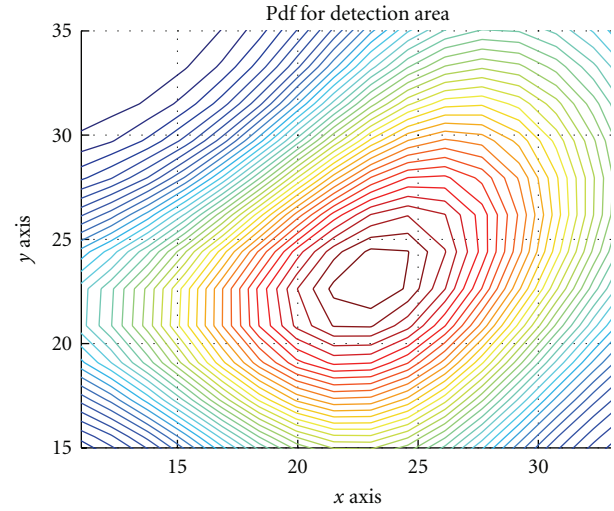


Figure 11: Corresponding 2-D contour mapping.

The result is given in Figure 9, marked as a gray rectangle. Compute the probability for each point in the black rectangle according to (20) and plot the 3-D probability contour in Figure 10. The corresponding 2D contour mapping is shown in Figure 11. From Figures 10 and 11, we chose the point with the largest possibility as the estimated PU location, that is, (22.4702, 22.9886). This estimation is really close to the actual location (23, 23).

The key of the algorithm is to get the precise estimation of the path loss coefficients following Step 4, as those estimations will be used to determine the PU location. The estimation accuracy depends on the reference point. The path loss for the reference point has been assumed to follow free-space propagation. Figure 12 shows how the actual path loss for the reference point affects the accuracy. It is shown that the effect is not that significant. We also check how the distance between the reference point and the PU location affects the accuracy, shown in Figure 13. It is shown that as the reference point is further away from the PU, the estimation gets worse. Consequently, the localization error follows the same trend, shown in Figure 14. The localization error increases with the reference distance. Thus, in order to achieve desired accuracy, the reference point should be close enough to the PU. For this specific example, the reference node should be chosen within 1.2 units away from the PU for good accuracy.

5. Conclusion

In this paper, we have proposed a solution to the hidden-node problem in a cooperative GPS enabled cognitive network. Our proposed localization algorithm estimates

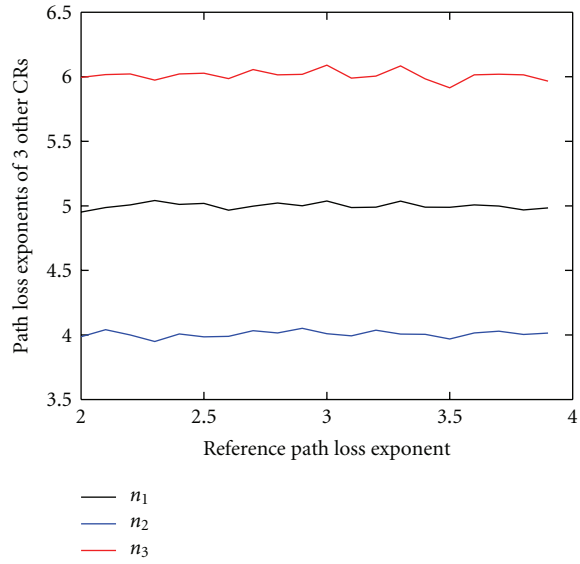


Figure 12: Estimated path loss of 3 CRs versus the actual path loss of reference point.

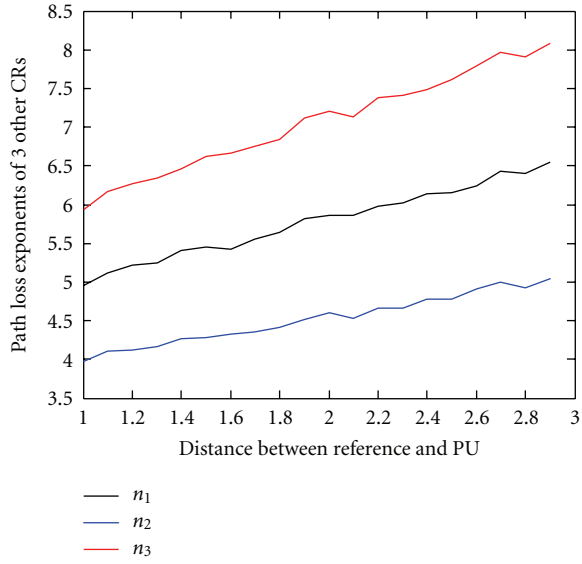


Figure 13: Estimated path loss of 3 CRs versus the distance between reference point and the PU location.

the path loss model and then locates the position of PU. Consequently, we can determine the disable region for CRs to avoid the interference to PU as well as the hidden-node problem. The simulation results have shown that the localization algorithm can provide accurate location results given that the reference node is close enough.

Acknowledgments

This work was supported by AFOSR under Grant no. FA9550-09-0630 and by NSF under Grant no. CNS 0619693.

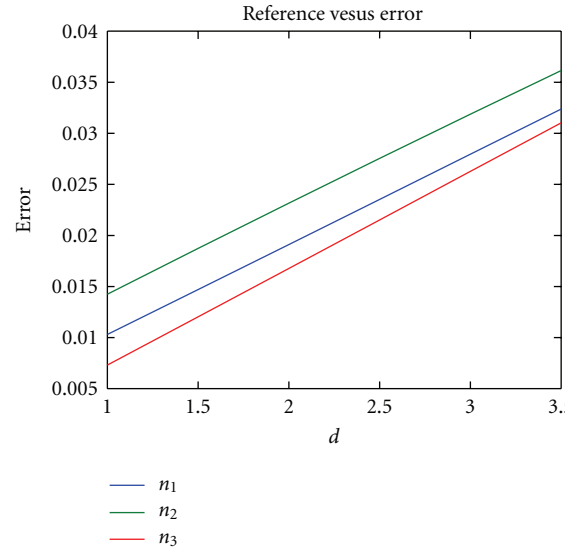


Figure 14: Algorithm error versus reference location.

References

[1] J. Mitola, *Cognitive radio*, licentiate thesis, KTH, Stsockholm, Sweden, 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] "Report FCC 05-57," ET Docket no. 03–108, March 2005.

[4] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proceedings of the 2005 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 131–136, Baltimore, Md, USA, November 2005.

[5] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 137–143, Baltimore, Md, USA, November 2005.

[6] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part II: multiuser networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2214–2222, 2007.

[7] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: two user networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2204–2212, 2007.

[8] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[9] F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: part II-the hidden terminal problem in carrier sense multiple-access and the busy-tone solution," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1417–1433, 1975.

[10] Z. J. Haas and J. Deng, "Dual busy tone multiple access (DBTMA)-a multiple access control scheme for ad hoc networks," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 975–985, 2002.

[11] G. Ganesan and Y. Li, "Agility improvement through cooperative diversity in cognitive radio," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, pp. 2505–2509, St. Louis, MO, usa, December 2005.

[12] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ANSI/IEEE Std 802.11: 1999 (E) Part 11, ISO/IEC 880211, 1999.

[13] A. H. A. Rahman and Z. A. Zukarnain, "Performance comparison of AODV, DSDV and I-DSDV routing protocols in mobile ad hoc networks," *European Journal of Scientific Research*, vol. 31, no. 4, pp. 566–576, 2009.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 1973.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 740594, 18 pages
doi:10.1155/2010/740594

Research Article

The Information Theoretic Approach to Signal Anomaly Detection for Cognitive Radio

Mostafa Afgani, Sinan Sinanović, and Harald Haas

The University of Edinburgh, AGB, King's Buildings, Mayfield Road, Edinburgh EH9 3JL, UK

Correspondence should be addressed to Mostafa Afgani, m.afgani@ed.ac.uk

Received 1 December 2009; Accepted 13 March 2010

Academic Editor: Massimiliano Laddomada

Efficient utilisation and sharing of limited spectrum resources in an autonomous fashion is one of the primary goals of cognitive radio. However, decentralised spectrum sharing can lead to interference scenarios that must be detected and characterised to help achieve the other goal of cognitive radio—reliable service for the end user. Interference events can be treated as unusual and therefore anomaly detection algorithms can be applied for their detection. Two complementary algorithms based on information theoretic measures of statistical distribution divergence and information content are proposed. The first method is applicable to signals with periodic structures and is based on the analysis of Kullback-Leibler divergence. The second utilises information content analysis to detect unusual events. Results from software and hardware implementations show that the proposed algorithms are effective, simple, and capable of processing high-speed signals in real time. Additionally, neither of the algorithms require demodulation of the signal.

1. Introduction

Cognitive radio (CR) is the term used to describe smart, reconfigurable wireless communications devices that are capable of automatically adjusting their operating characteristics in order to adapt to changes in the radio environment. The purpose of such a system is to enable efficient use of the available radio spectrum and provide reliable service to the end user [2]. The motivation for efficient spectrum utilisation arises from the fact that it is a very limited resource. Although the electromagnetic spectrum is (for all intents and purposes) infinite, only a small fraction of it is usable for personal wireless communications as we know it today. Furthermore, while the spectrum available remains fixed, the number of wide-band wireless systems contending for access keeps growing—further compounding the spectrum scarcity problem.

Traditionally, the radio spectrum has been divided into a number of usable bands by regulatory bodies such as the Federal Communications Commission (FCC) in the USA and the Office of Communications (Ofcom) in the UK. Each of the bands is then assigned for exclusive access by a particular operator or service. A notable exception is of course the set of bands known as the industrial, scientific and medical (ISM) bands where emission from unlicensed consumer electronic devices is tolerated. While this restrictive approach to sharing the radio spectrum succeeds at providing a certain degree of interference protection, it is an inefficient use of the available resources since it is extremely unlikely that all of the bands are in use at the same time at a given place.

CR systems aim to simultaneously provide better quality of service and spectrum utilisation by dynamically moving the communication link from crowded or occupied bands to ones that do not appear to be in use by a primary licensed system at that instant. In order to carry out this task, secondary CR devices perform *spectrum sensing*—a procedure used to identify "holes" (free bands) in the spectrum and characterise the radio environment [3]. While there are a number of diverse approaches to problem [4], none of them are perfect. Energy detection-based methods [5] are limited by signal-to-noise ratio (SNR) constraints while methods relying on cyclostationary features [6] are limited by the amount of a priori information available regarding the signal structure of the primary system. As a result of these shortcomings, spectrum sensing cannot completely avert the risk of interference that arises from

dynamic spectrum sharing. Since interference generally leads to *anomalous* signal behaviour, an additional layer of simple signal processing algorithms that can help detect and characterise that behaviour is useful.

Anomaly detection refers to the process of locating unusual and unexpected events that may exist alongside nominal samples in a dataset. It is a process that is already utilised in a large number of diverse application domains. Typical examples include the detection of: unauthorised access to computer systems [7], irregularities in vital signs such as electrocardiogram (ECG) traces [8], fraud in financial services [9], and so forth. An extensive survey of current anomaly detection techniques and application domains is provided in [10].

The aforementioned survey reveals that there are many different approaches to solving the anomaly detection problem—each with its own set of advantages and disadvantages. However, there is one drawback that is shared by most algorithms: computational complexity. The computational effort required makes it difficult to adapt these techniques for real time and online processing of the input signal. This is unfortunate since any algorithm employed on an interactive communications system such as a CR platform must be capable of real time operation to maintain a seamless user experience. To overcome this challenge, two complementary anomaly detection algorithms based on simple information theoretic measures have been developed and are presented in this paper. The first method utilises Kullback-Leibler divergence (KLD) [11] while the latter uses the information content of individual signal events [12]. The algorithms are easy to generalise and broadly define anomalies as events that lead to changes in the nominal probability distribution of the radio signal. As a result, it is possible to employ the techniques for the detection of a wide range of disruptive events such as interference, timing errors, transmitter malfunction, and so on.

KLD is a convenient and robust method of measuring the difference between two data sets in a statistical sense. Due to its versatility and general appeal, it finds use in fields as diverse as economics [13] and computational neuroscience [14]. As a statistical comparison tool, KLD can also be employed for the automatic and real time detection of unusual (anomalous) data segments. The proposed KLD-based technique utilises two data windows to perform a statistical comparison of neighbouring segments of signals with periodic structures (e.g., systems utilising time division multiple access [15]). Since segments separated by the signal period are expected to be analogous and hence have similar statistical characteristics, any deviation can be taken to imply the presence of an anomaly.

Unlike the KLD-based method, the information content analysis (ICA) algorithm can also be applied to signals lacking any kind of periodic features. Information content is a quantity that is directly related to the probability of an event: the lower the probability, the higher the information content. Since anomalies are, by definition, rare (low probability), the associated information content is high. The proposed anomaly detection algorithm exploits this fact by analysing the signal for high-information content events.

Software implementations of the algorithms have been tested against a set of real wireless signals with promising results. Additionally, a Xilinx Virtex4 field-programmable gate array-(FPGA-) based hardware implementation of the KLD-based method has shown that the algorithm is indeed capable of real time analysis of high speed, high bandwidth signals.

A brief review of some of the anomaly detection algorithms described in literature is provided in Section 2 while the proposed algorithms are described in detail in Section 3. Results from applying the techniques to the test signals and measures of performance are provided in Section 4. The hardware implementation is briefly discussed in Section 5 while Section 6 concludes the paper with a summary of the contributions made and directions for future work.

2. Review of Existing Methods

Anomaly detection, also known as novelty detection or outlier detection, is a rich field of research with a very large body of work that exists in the literature. The existence of multiple survey-type papers such as [10, 16–20] is a testament to the true extent of the subject of anomaly detection. It is therefore surprising to learn that it is still very much an active area of research lacking generic algorithms that can be applied universally to anomaly detection problems. Most of the methods described in literature are based on tightly constrained frameworks that apply to very specific classes of problems.

Existing techniques of anomaly detection can be separated into a handful of classes depending on the underlying approach. *Classification*-based methods utilise supervised machine learning techniques to categorise nominal and anomalous behaviour while *clustering* and *nearest-neighbour* based techniques rely on measures of the relative distance between points of data. *Statistical* techniques detect anomalies by comparing the test data points against stochastic models of nominal behaviour. *Information theoretic* methods employ measures of information such as Kolmogorov complexity and entropy and work under the assumption that anomalies lead to a change in the information content. The algorithms proposed in this paper employ techniques that are both statistical and information theoretic in nature.

A statistical method of detecting anomalies in sensor data streams is proposed by Basu and Meckesheimer in [21]. Relying on the assumption that the data stream is continuous, the method exploits the fact that correlation between neighbouring data points is higher than between points separated by a relatively long length of time. The described algorithm detects anomalous events by comparing the value of each event against the median of a data set composed of neighbouring events. The performance of the method then depends on the size of the data set and the threshold. Since the algorithm expects an input where subsequent data points change little under nominal circumstances, it is unsuitable for use in typical communications systems where the signal strength can vary considerably even under normal operating conditions.

An algorithm for detecting anomalous network traffic by means of a combined statistical and information theoretic

measure is described by Krügel et al. in [7]. For each packet, an anomaly score is computed by considering the packet type, length, and payload distribution. If the combined score exceeds a certain threshold established through training, existence of an anomalous packet is signalled. Since the algorithm is designed for operation in the network layer, it cannot be utilised for link monitoring and anomaly detection in the physical layer.

Another set of statistical anomaly detection algorithms are presented by Desforges et al. and Yeung and Chow in [22, 23], respectively. Both papers propose the use of the Parzen windows method of nonparametric smooth probability density estimation in order to establish a stochastic model of the data distribution. While Yeung and Chow simply test whether a data point belongs to a given model, Desforges et al. also construct a model of the test data set and compare that against the reference. Since the model of the underlying process is determined once at the onset of the experiments, the algorithms cannot cope with nonstationary systems. Utilisation of Parzen windows method for density estimation also makes the algorithms computationally expensive and unsuitable for real time implementation.

A technique for detecting anomalous segments ("discords") in structured time series such as ECG traces is described by Lin et al. in [8]. Given a time series containing a discord, the algorithm essentially splits the series into a set of small segments and computes the mutual distance between the segments. If a segment is then found to have a minimum distance larger than a predefined threshold, it is labelled as anomalous. Although the algorithm shows promising results, it is unsuitable for real time implementation due to the computational complexity cost associated with performing a search for anomalous segments.

Finally, the use of various information theoretic measures for anomaly detection is discussed by Lee and Xiang in [24]. However, the focus of the paper is on determining the suitability of data models through the use of measures such as entropy and relative entropy (i.e., KLD) rather than algorithms for detecting anomalies.

It is evident from this survey of existing techniques that there is a lack of algorithms that offer the features needed (nonparametric with a low computational complexity and the ability to handle nonstationary behaviour) to analyse radio frequency signal envelopes in real time for anomalies.

3. Anomaly Detection Algorithms

The detection algorithms utilise KLD and information content analysis, respectively, to determine the presence of anomalies. Both quantities are ultimately calculated from estimates of the statistical probabilities of events in the signal.

Given two data sets P_n and Q_n, at time n, that contain samples from domain X, it is possible to obtain empirical estimates of the associated probability mass functions (PMFs) p_n and q_n from a nonparametric model such as a histogram. Once the PMF estimates are available, the KLD between them can be calculated using [11]

$$D(p_n \| q_n) = \sum_{x \in X} p_n(x) \log_2 \frac{p_n(x)}{q_n(x)}, \tag{1}$$

where $x \in X$. Since base-2 logarithm is used, the divergence is measured in *bits*. KLD between two PMFs is generally asymmetric: that is, $D(p_n \| q_n) \neq D(q_n \| p_n)$ and the triangle inequality is not satisfied. When $p_n = q_n$, the KLD is zero; otherwise, it is a positive real number ($\mathbb{R}_+$). For brevity and convenience, $D(p_n \| q_n)$ will also be referred to as D_n in this paper.

KLD belongs to a class of distance measures known as *f-divergence* (or *Ali-Silvey distances*). Some of the other distance measures that belong to the same class are variational distance (symmetric), Hellinger distance (symmetric), and Chernoff distance (generally asymmetric) [25]. While they are all equally suitable for quantifying the statistical difference between two probability distributions, KLD and variational distance are the least complex and therefore the easiest to implement. Variational distance is defined as

$$\begin{aligned} V(p_n \| q_n) &= \frac{1}{2} \sum_{x \in X} | \, p_n(x) - q_n(x) \, | \\ &= \frac{1}{2} || p_n - q_n ||_1, \end{aligned} \tag{2}$$

where $\| p_n - q_n \|_1$ is commonly known as the $\mathcal{L}_1$ distance (L1D) between the PMFs p_n and q_n. Furthermore, KLD and L1D (and hence the variational distance) are related by the inequality [11]

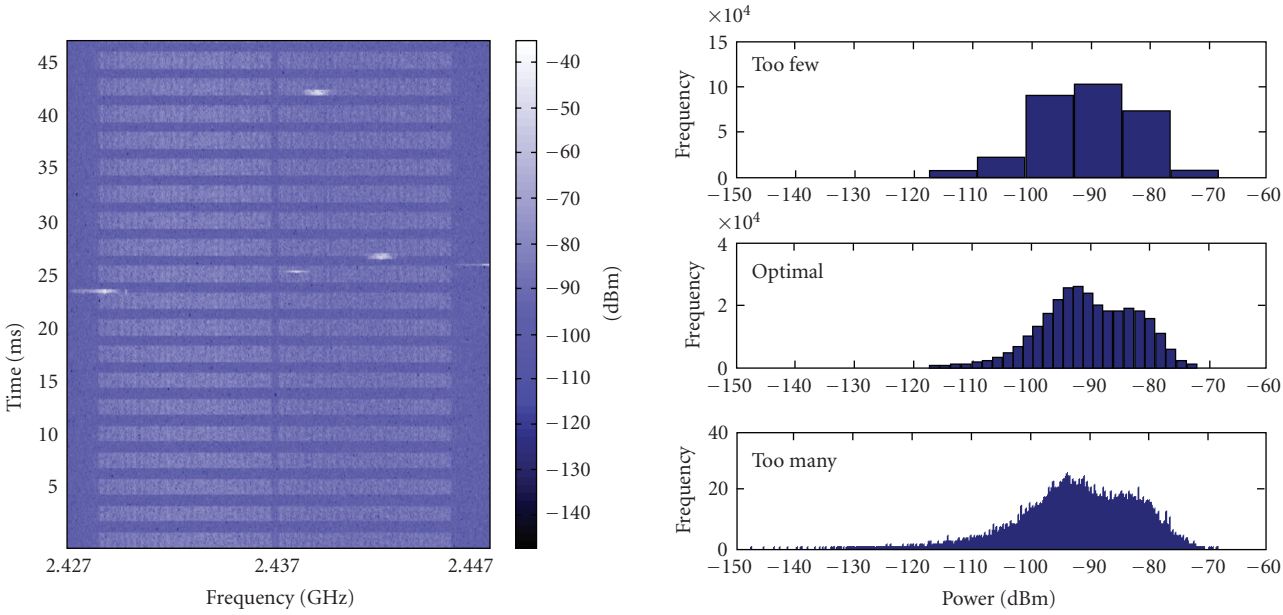
$$D(p_n \| q_n) \geq \frac{1}{2 \ln 2} || p_n - q_n ||_1^2. \tag{3}$$

Crucially, it states that $D(p_n \| q_n)$ is bounded by $\| p_n - q_n \|_1^2$ and not $\| p_n - q_n \|_1$. It is an important distinction as it implies that for certain PMF pairs the KLD may in fact be *smaller* than the L1D. For a pair of largely dissimilar PMFs (Differences that are large enough to produce a $\mathcal{L}_1$ distance of $2 \ln 2$ or greater, to be precise.), as is generally the case when comparing an anomalous data set against a nominal reference, larger distance magnitudes are obtained from KLD rather than L1D. However, when both PMFs are similar (e.g., a nominal data set and the reference), this can lead to L1D values that are larger compared to KLD—increasing the likelihood that false positives are detected. As a result, it is expected that KLD is better suited for statistical anomaly detection compared to L1D. This is confirmed by the results seen in Section 4 where the performance of a KLD-based algorithm is compared against one based on L1D. The algorithm for anomaly detection using KLD is described in Section 3.2.

Information content analysis is another technique based on an information theoretic quantity that can be utilised for the detection of anomalies. The amount of information, $I_n(x)$, conveyed by any discrete random event, x_n, at time n, is directly related to its probability of occurrence, $p_n(x)$ [12]:

$$I_n(x) = -\log_2 \{ p_n(x) \}, \quad n = 1, 2, \ldots \tag{4}$$

Since base-2 logarithm is used once again, information is also measured in *bits*. The equation implies that an event with a very high probability of occurrence carries very little information while a large amount of information is



(a) Signal with interference

(b) Effect of bin count

FIGURE 1: (a) Spectrogram of a wireless local area network (WLAN) signal experiencing interference from a Bluetooth device. The regular pattern is a single WLAN frame repeatedly transmitted by a signal generator. The frequency hopping nature of the Bluetooth transmission is clearly visible in the plot. (b) Impact of the number of bins utilised, β, on a histogram of the instantaneous power density of the WLAN signal. When β is too small, the histogram is insensitive to small changes and does not effectively capture the subtleties of the process. At the other extreme, it is too sensitive and therefore susceptible to noise. The optimal β is with respect to some minimum error criterion [1] and provides a good balance between resolution and sensitivity.

conveyed by the occurrence of rare events (i.e., $I_n(x) \to \infty$ as $p_n(x) \to 0$). Information is always real, positive ($\mathbb{R}_+$) and monotonically increasing with decreasing values of event probability. ICA is essentially a nonlinear scaling function that favours the unusual.

3.1. Histogram and PMF Estimation. It is clear from (1) and (4) that both KLD analysis and ICA require estimation of empirical event probabilities. One approach to obtaining the necessary estimates is via event histograms. In addition to being simple to implement, histograms are nonparametric—implying that no assumptions need to be made regarding the underlying distribution of the sample data.

For samples that originate from domain X, the histogram is obtained by first partitioning X into bins B such that

$$X = \bigcup_{l=1}^{\beta} B_l, \tag{5}$$

and then counting the number of samples that belong to each bin. β is the total number of bins used to construct the histogram. Once the histogram is available, the empirical PMF of the sample set is easily obtained by simply dividing the histogram by the cardinality of the set.

Given a statistically significant sample size, it is clear that the only parameter that affects the quality of the PMF estimate obtained is the bin allocation B. If the partitions are then assumed to be equidistant for simplicity, the only variable that remains is the number of bins utilised: β. The effect of β on the histogram of a random process is shown in Figure 1. The random process in question is the instantaneous power density at any time-frequency point of the signal shown in Figure 1(a). It is a wireless local area network (WLAN) signal experiencing bursts of interference from a Bluetooth (BT) device. Histograms of the power density obtained using three different values of β are shown in Figure 1(b).

When a small number of bins are utilised, that is, β is small, the histogram is insensitive to small scale variations in the input. As a result of the poor resolution, the estimated model fails to adequately capture the subtleties in the behaviour of the underlying random process. On the other hand, when the value of β utilised is too large, the resolution is too high and the histogram is overly sensitive—resulting in an estimate that is noisy. The optimal value of β yields a good balance between resolution and sensitivity.

A method of computing the optimal bin size (and hence the optimal β) for constructing a histogram, subject to some minimum mean square error criterion, is provided in [1]. While the algorithm described therein is conceptually simple, it unfortunately requires the use of exhaustive search to iteratively minimise a certain cost function—making it too computationally expensive to be evaluated in real time on a hand-held mobile device with limited energy and processing power.

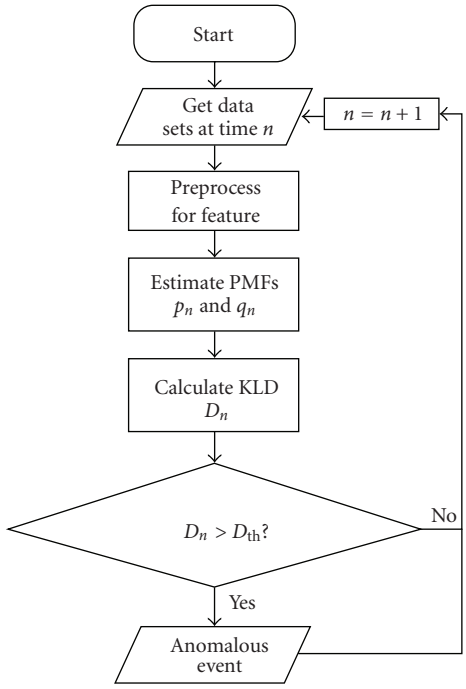


Figure 2: The algorithmic flowchart. KLD is used to compare the statistical distribution of a test data set against that of a reference. If the divergence, D_n, is greater than some predefined threshold, D_{th}, the test set may be anomalous.

The impact of β on the performance of each of the anomaly detection algorithms has been investigated and the results are presented in Section 4.2.4.

Depending on the choice of B, there may be zeros in the estimated PMFs due to the presence of empty histogram bins. Under such circumstances, calculation of the KLD using (1) can be a problem as it leads to instances where $0\log_2\{0/q_n(x)\}$ or $p_n(x)\log_2\{p_n(x)/0\}$ have to be evaluated. While it is certainly possible to handle these as special cases by setting them to 0 and, ∞, respectively, through continuity arguments, it may be better to simply avoid zeros in the PMFs. It is possible to avoid empty histogram bins and hence zeros in PMFs by adding a small number, λ, to every bin of the histogram. As preloading of histogram bins in this manner undoubtedly distorts the estimate of the true PMF, the preload value must be carefully chosen. According to the work done by Krichevsky and Trofimov [26] and Johnson et al. [27] $\lambda = 0.5$ is a good choice.

3.2. Algorithm Based on KLD. The capability of KLD to quantise the difference, in a statistical sense, between two data sets to single real value is ideal for use in anomaly detection since it provides a convenient detection metric. A general description of the algorithm is provided here while a discussion of the optimisations needed for an efficient hardware implementation is provided in Section 3.3 that follows.

The flowchart in Figure 2 shows the proposed algorithm. At time n, the process starts with the acquisition of the two data sets to be compared using KLD. One of the data sets is a reference (Q_n) while the other is the one under test (P_n). If the samples in the data sets do not directly represent the parameter of interest, they must be processed. Once the data sets have been suitably transformed, the associated PMFs $p_n(x)$ and $q_n(x)$ are estimated and used to compute the KLD, D_n. If D_n is then observed to be larger than some predefined KLD threshold, D_{th}, the test data set may be anomalous.

Table 1: Complexity analysis of KLD.

Operations	Σ	$\times$	$\div$	log	Total
	$2\lvert P_n\rvert + \beta$	β	3β	β	$2\lvert P_n\rvert + 6\beta$
Memory	$\lvert P_n\rvert + \psi + 2\beta$				

This general approach to detecting anomalies using KLD can be easily adapted for use with signals containing periodic structures. One example of such a signal is IEEE 802.16e wireless broadband (WiBro) which utilises time division duplexing (TDD) [28]. Periodic signals are expected to have statistics that are also periodic—implying that segments of the signal separated by the period, T_p, should have probability distributions that are very similar under normal circumstances. Therefore, by simply acquiring the data sets P_n and Q_n from two sliding signal windows of length T_w with centres separated by T_p, the proposed algorithm can be utilised for the detection of anomalies in periodic signals. KLD analysis can be performed on the signal envelope itself and as a result, demodulation is unnecessary and the only a priori information required by the algorithm is the signal period T_p.

While the steps required to compute the KLD are all simple and straightforward, the storage (data buffers) and the number of arithmetic operations required grow linearly with the size of input data sets. As these data sets can be very large when analysing high speed signals, it can easily lead to scenarios where it may not be possible to provide for the resources required by the algorithm. Analysis of the algorithm's complexity and memory requirements follows and is summarised in Table 1.

The input data sets P_n and Q_n themselves require a buffer capable of holding at least $|P_n| + \psi$ elements, where ψ is the number of samples corresponding to the signal period and $|P_n|(= |Q_n|)$ is the size of the data windows in samples. Only a single buffer is required for the input data since one of the data sets is essentially just a ψ-delayed version of the other in this case. Computing the frequency count over the bins (B), for the purpose of estimating the histograms, requires up to $|P_n|$ additions for each of the two windows. Once the histograms are available, the PMFs are obtained by dividing the frequency count in each of the β bins by $|P_n|$. Two buffers of size β each are then required to store the resulting PMFs. Computation of the KLD from the PMFs then requires a further β divisions, logarithms, multiplication, and addition, respectively.

3.3. Hardware Implementation. The analysis performed in the previous Section 3.2 reveals that a direct interpretation of the algorithm to hardware would be inefficient, inflexible,

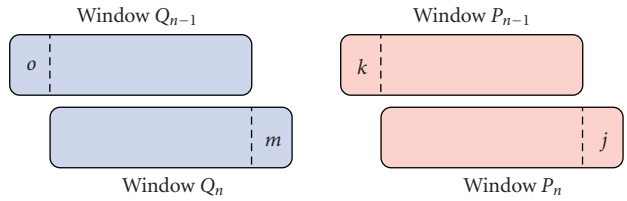


Figure 3: At any given time, only a maximum of four histogram bins need to be updated (two per window). Bins incremented (fresh samples) are denoted by j and m while bins decremented (old samples at the end of the window) are denoted by k and o.

Table 2: Complexity analysis of DKLD.

Operations	Σ	×	÷	log	Total
	20	8	16	16	60
Memory	$\lvert P_n\rvert + \psi + 2\beta$				

and computationally expensive. The inefficiency arises from the fact that at each successive time instance, the PMFs and the KLD are completely recalculated, even though it is only a single sample that changes in each of the data sets. The inflexibility comes from the fact that the computational complexity depends on β, implying that a direct interpretation would be limited by the initial choice of the histogram bin resolution. Finally, logarithms and divisions can be very costly to implement in hardware. Fortunately, there are several well-known methods that can be adopted to overcome each of these challenges.

The complexity that arises from the division operation in (1) can be removed by exploiting the identity $\log(a/b) = \log(a) - \log(b)$:

$$\begin{aligned} D_n &= \sum_{x\in X} p_n(x)\log_2\frac{p_n(x)}{q_n(x)} \\ &= \sum_{x\in X} p_n(x)\Big\{\log_2[p_n(x)] - \log_2[q_n(x)]\Big\}. \end{aligned} \tag{6}$$

The division operation is exchanged for a subtraction and a second base-2 logarithm operation which can be implemented in a very efficient manner by means of a lookup table.

Further efficiency improvements can be achieved by making application-specific changes to the way the algorithm is evaluated. Since the purpose of the algorithm is to analyse periodic signals by means of two sliding windows, it holds that at any given instance, only one sample in each of the data sets changes. This in turn implies that only a maximum of 4 histogram/PMF bins need to be updated at that instant—two for each data set/window. The two bins per window account for the freshly acquired sample (bin frequency count incremented by one) and the sample that is dropped at the end of the window (bin frequency count reduced by one). An illustration is provided in Figure 3.

This also means that the KLD values change very little between subsequent time steps for this particular application—suggesting that it is possible to rewrite (6) in the form of a differential equation:

$$\begin{aligned} D_n &= \sum_{x\in X} p_n(x)\Big\{\log_2[p_n(x)] - \log_2[q_n(x)]\Big\} \\ &= D_{n-1} \\ &\quad - p_{n-1}(j)\Big\{\log_2[p_{n-1}(j)] - \log_2[q_{n-1}(j)]\Big\} \\ &\quad + p_n(j)\Big\{\log_2[p_n(j)] - \log_2[q_n(j)]\Big\} \\ &\quad - p_{n-1}(k)\Big\{\log_2[p_{n-1}(k)] - \log_2[q_{n-1}(k)]\Big\} \\ &\quad + p_n(k)\Big\{\log_2[p_n(k)] - \log_2[q_n(k)]\Big\} \\ &\quad - p_{n-1}(m)\Big\{\log_2[p_{n-1}(m)] - \log_2[q_{n-1}(m)]\Big\} \\ &\quad + p_n(m)\Big\{\log_2[p_n(m)] - \log_2[q_n(m)]\Big\} \\ &\quad - p_{n-1}(o)\Big\{\log_2[p_{n-1}(o)] - \log_2[q_{n-1}(o)]\Big\} \\ &\quad + p_n(o)\Big\{\log_2[p_n(o)] - \log_2[q_n(o)]\Big\}, \end{aligned} \tag{7}$$

where the four bin indices j, k, m, and o are assumed to be unique. If not, any duplicate terms in the equation are set to zero.

The differential equation form of KLD (DKLD) shows that its computational complexity is no longer dependent on the number of histogram bins utilised in evaluating the PMFs. Assuming that D_{n-1} is available, only 16 additions/subtractions, 16 logarithms, and 8 multiplications are needed to calculate D_n—regardless of the value of β. This opens the path for a fixed complexity, flexible, and efficient implementation that can be easily updated to accommodate a wide range of histogram resolutions.

The computational complexity and storage requirement of DKLD are shown in Table 2. Comparisons against the unmodified, direct interpretation version of KLD (Table 1) reveals that while memory utilisation remains unchanged, there is a vast difference in the number of operations required. Regardless of the window size and histogram bin count, 60 operations are needed to compute the KLD. In addition to the 16 additions/subtractions, 16 logarithms, and 8 multiplications required for the DKLD (7), 4 more additions/subtractions are required to update the affected histogram bins and 16 divisions are required to obtain the necessary PMFs at times $n-1$ and n from the histogram.

Switching to a fixed-point representation and using a lookup table for base-2 logarithms provide further reductions in complexity at the expense of a slight increase in the memory requirements. The size of the table, L, then dictates the precision available. Additionally, if $\lvert P_n\rvert$ is chosen such that it is always a power of two (PoT), that is,

$$\lvert P_n\rvert = 2^{\sigma}, \quad \sigma = 0, 1, 2, \ldots, \tag{8}$$

no division operations are required to obtain the PMFs since division by a PoT is simply a bit-shift operation that costs nothing in hardware.

Table 3: Complexity analysis of FP-DKLD.

Operations	Σ	×	÷	log	Total
	20	8	0	0	28
Memory	$\lvert P_n\rvert + \psi + 2\beta + L$				

The complexity and storage requirements of a fixed-point DKLD (FP-DKLD) based algorithm utilising a log lookup table and PoT constraint on $|P_n|$ is shown in Table 3. It can be seen that with some simple changes and constraints, the complexity of the anomaly detection algorithm can be greatly reduced—allowing for efficient and high speed hardware implementations. Results obtained from a Xilinx Virtex4 FPGA implementation of the FP-DKLD algorithm are shown in Section 5.

3.4. Information Content Analysis Algorithm. Unlike the KLD-based anomaly detection algorithm just described, the ICA-based method analyses individual input samples rather than aggregate sets of data. The information conveyed by the events is the detection metric utilised. The following is a general description of the algorithm.

First and foremost, it is necessary to establish the type of event that is under observation. This can be any property that is associated with the signal under test (e.g., instantaneous amplitude, phase, or power). If the event type chosen is measurable directly from the signal envelope, demodulation is unnecessary for anomaly detection. The ICA algorithm utilises supervised learning to establish a reference histogram (and hence probability) of events; therefore, some clean signal is required for training. Once the reference histogram is obtained, online analysis of the signal under test can commence. Events from the test signal are extracted and used to update the reference histogram. This yields updated event probabilities and hence the associated information content. If the information content $I_n(x)$ of any event x_n, at time n, is above some predefined threshold I_{th}, an anomaly may be present.

Once again, it is clear that the event histogram plays a central role in the anomaly detection algorithm. It has been stated previously in Section 3.1 and illustrated by Figure 1 that the number of bins utilised, β, has a significant impact on the sensitivity of the histogram and hence the effectiveness of the detection algorithms. When β is too small, anomalous events may not be detected due to poor sensitivity—leading to missed detections. On the other hand, when β is too large, even nominal events will appear to have low probability—leading to a large number of false positives. It is therefore necessary to find a β that offers a good balance between sensitivity and probability of detecting false positives.

The event histograms shown in Figure 1(b) reveal another potential challenge for the ICA algorithm. It can be seen that the histograms have long tails with numerous low probability (i.e., high information content) events even when the signal is behaving nominally. Although this is expected for any analogue signal transmitted over a lossy physical channel, it raises the possibility that numerous false positives are observed at a detector that employs a simple information content threshold. It is certainly possible to reduce the number of tail events by using a smaller number of bins, but that leads to reduction in sensitivity and hence an increase in the probability of missed detections.

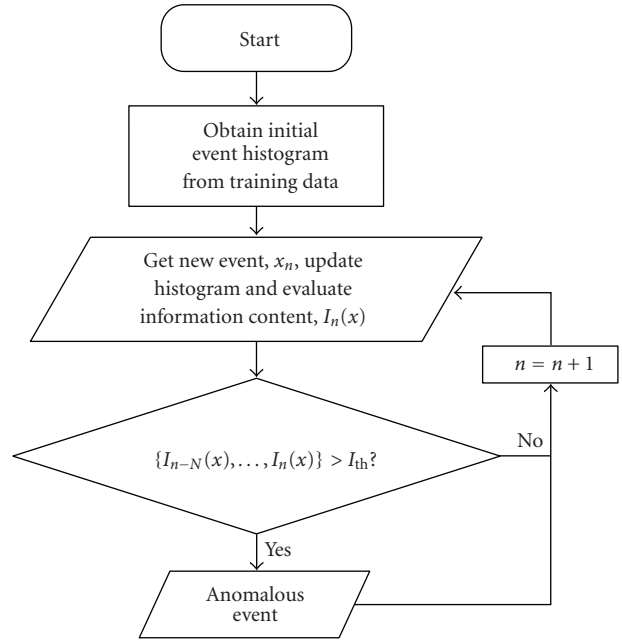


Figure 4: Flowchart of information content analysis algorithm with clustered anomaly detection. First, clean data is used to initialise the reference event histogram and event probabilities. Then, events in the signal under test are used to update the reference histogram and event probabilities. The updated values are used to estimate the information content of the events. If the information content of $N+1$ contiguous events exceeds a predefined threshold, I_{th}, an anomaly may be present in the signal under test.

Examination of the interference scenario in Figure 1(a) reveals an important distinction between anomalous events and the underlying signal—anomalies tend to appear in clusters while nominal low-probability signal events are decidedly "singular." This difference is the key to reducing the number of false positives while still maintaining a low rate of missed detections. The proposed algorithm is easily augmented to benefit from this insight: instead of triggering on individual high information content events, the detector must search for contiguous groups of events that exceed the predefined information content threshold.

A flowchart of the algorithm with simplified clustering is shown in Figure 4. The general approach is as before, with the exception of the last step. With the simple clustering extension, detection of an anomaly is signalled only when a contiguous sequence of N previous events and the current event exceeds a predefined information content threshold. Sequence detection is used rather than full two-dimensional clustering to minimise the complexity of the algorithm. This is permissible since a sequence can be considered as a one-dimensional cluster. The effect of the cluster size utilised on the detector performance is examined in Section 4.2.5.

Table 4: Runtime complexity analysis of ICA.

Operations	Σ	×	÷	log	Total
	2	0	1	1	4
Memory	$\beta + N + 2$				

The discrimination threshold is an important aspect of any detector. While the optimum threshold is problem and cost function specific, it is generally chosen to minimise missed detections while still maintaining a low rate of false positives. For the proposed anomaly detection algorithm, it is not possible to define a single information content threshold, I_{th}, that is suitable for use with any arbitrary signal. I_{th} is signal specific and may be set automatically using information obtained from the clean training data. After the reference event histogram and probabilities have been estimated, the reference information content associated with each event type can be easily computed using (4). For β bins, the standard deviation, $\sigma_{\text{I}\langle\beta\rangle}$, of the reference information content provides a measure of the spread and may be used to obtain the threshold:

$$I_{\text{th}} = m\sigma_{\text{I}\langle\beta\rangle}. \tag{9}$$

m is a multiplicative factor greater than 1. The effect of I_{th} on detector performance is investigated in Section 4.2.3.

Due to the simplicity of the ICA algorithm, its runtime operational complexity and memory requirements are negligibly small. On completion of the initial training phase, a small buffer capable of holding just $\beta + 1$ elements is required to store the event histogram and the total events count. At runtime, analysis of an event requires 2 additions to increment the relevant bin count and the total events count. Division of the incremented bin count by the total is then needed to obtain the event probability. After the probability is computed, a single base-2 logarithm is needed to calculate the event's information content. An additional buffer capable of holding $N + 1$ elements is also needed to accommodate information content clustering. A summery of the complexity analysis is provided in Table 4. It reveals that in addition to being negligibly small, the fixed runtime operational complexity is independent of any algorithmic parameter (e.g., histogram resolution)—suggesting that fast and efficient implementations for power limited hand-held devices are possible.

4. Results

In order to evaluate the performance of the proposed anomaly detection schemes, signals with different classes of abnormalities are employed as test cases. All of the signals under test are actual radio frequency transmissions captured using spectrum analysis hardware and therefore represent scenarios likely to be encountered by real world wireless devices.

Analyses of the test signals are provided in the following section while detailed performance analyses of the algorithms based on parameters such as histogram bin resolution, data window size, sampling rate and cluster length are provided in Section 4.2.

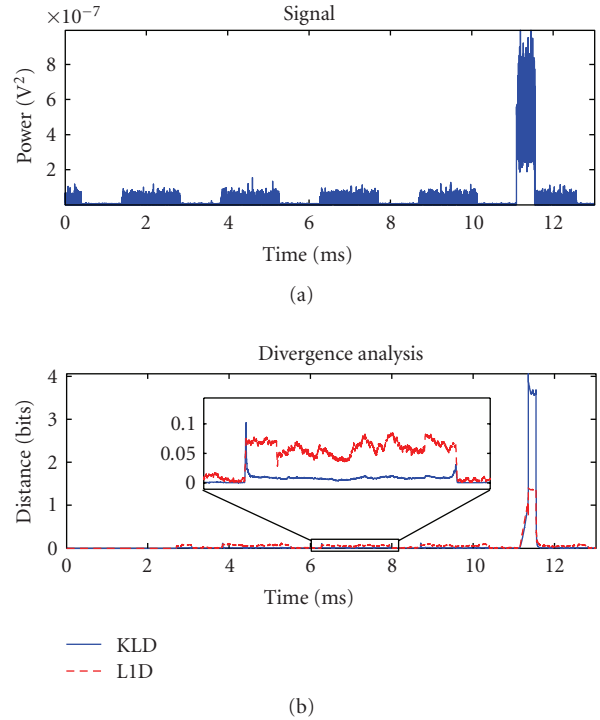


Figure 5: (a) WLAN signal with interference at 11.5 ms from a Bluetooth device. (b) Both KLD and L1D analyses of the signal result in detection of the anomaly. KLD appears to be better than L1D since it leads to a much larger peak and lower noise.

4.1. Data Analysis. Of the four data sets available, the first three are used to demonstrate the PMF divergence analysis (KLD/L1D)-based technique while the last is used to demonstrate the ICA-based technique.

4.1.1. Test Signal A. The signal is shown as a time series in Figure 5(a). It consists of a single WLAN frame that repeats with a period of 2.45 ms and a burst of interference from a Bluetooth device that is visible at 11.5 ms. The signal is similar to that shown earlier in Figure 1.

Both KLD and L1D are used to analyse the signal for the purpose of obtaining results that can be directly compared. Two windows with a duration of 256 μs each are employed to process the time series signal. The window centres are separated by 2.45 ms to match the WLAN frame repetition interval. The windows estimate the PMFs of the signal power. The number of histogram bins utilised is the optimal value (51 in this case) as obtained from the algorithm proposed by Shimazaki and Shinomoto [1]. In any case, it is shown in a subsequent Section (4.2.4) that the number of histogram bins used does not have a significant impact on the outcome—therefore, an arbitrary but reasonable choice such as 32 can also be used instead.

The result of the analysis is also shown in Figure 5. Both KLD-and L1D-based methods are successful at detecting

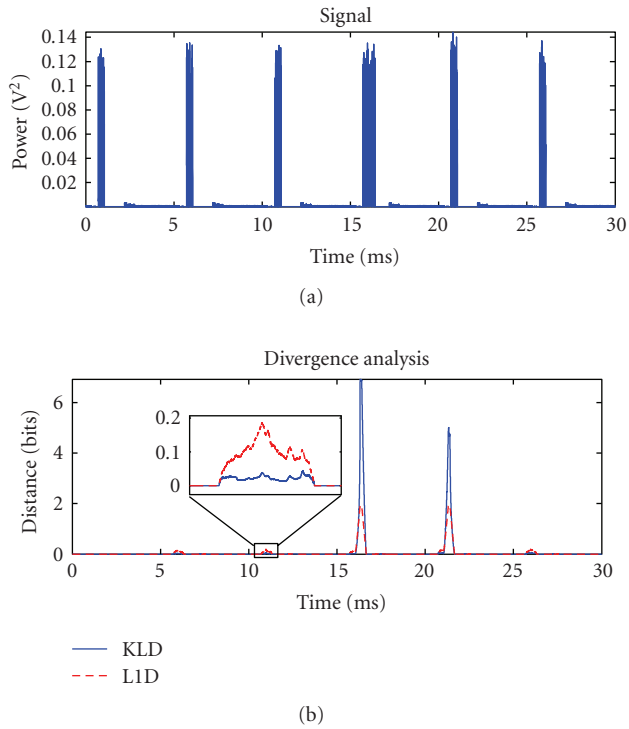


Figure 6: (a) Wireless broadband (WiBro) signal. The uplink subframe at 16 ms has a longer duration than others and is therefore unusual in this context. (b) KLD and L1D analysis both reveal the anomalous segment of the WiBro signal. A second peak is obtained when the signal returns to normal—this is due to the twin-windowing nature of the anomaly detection algorithm. Once again, KLD analysis results in a larger peak and lower noise.

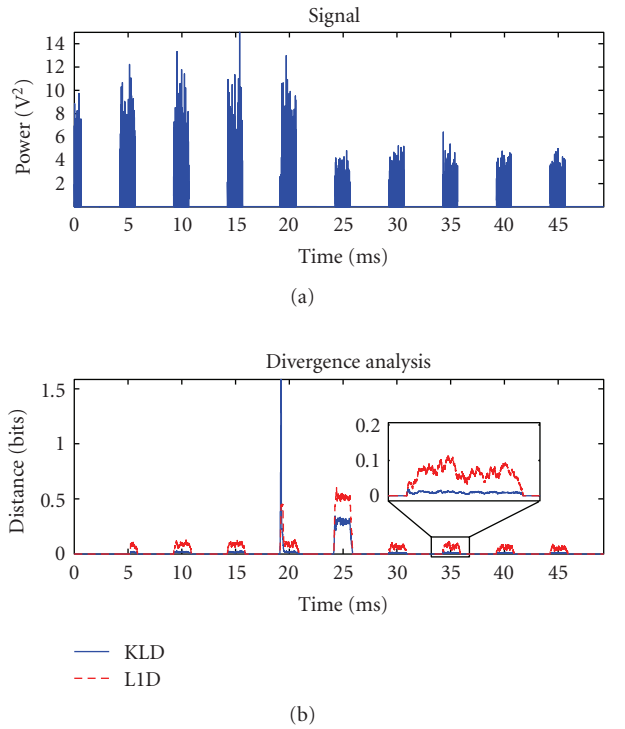


Figure 7: (a) Message exchange between a mobile terminal and a base station using the WiBro communications protocol. An extra command sequence in the uplink subframe at 20 ms initiates the power control loop. (b) The extra command sequence (20 ms) and subsequent power change (25 ms) are both revealed by KLD/L1D analyses of the signal. A larger KLD peak results from the anomalous command but not from the power change. L1D also leads to larger noise levels compared to KLD.

the presence of the anomaly (BT interferer). However, it is clear that KLD is the better choice as it produces a larger peak compared to L1D when the anomaly is detected. The baseline noise level with KLD is also much lower than that obtained with L1D—confirming the hypothesis presented in Section 3.

4.1.2. Test Signal B. The second signal under test is a wireless broadband (WiBro) signal. It is shown in Figure 6(a). Due to the proximity of the recording equipment to the mobile terminal (MT), the uplink (UL) subframes show a higher power level then downlink (DL) subframes. From the plot, it can be seen that the UL subframe at 16 ms is longer than any of the other UL subframes. In context of this particular signal snapshot, this behaviour is unusual and hence can be considered to be anomalous. Once again, two windows with a duration of 256 μs each are employed to estimate the signal power PMFs. The window centres are separated by 5 ms—corresponding to the frame period of the signal. The optimal histogram bin allocation scheme in [1] is once again used to determine the number of bins utilised ($\beta = 51$).

Analysis of the signal is also shown in Figure 6. A sharp peak in the divergence at 16 ms reveals the presence of the unusual UL subframe. A second peak is obtained when the signal returns to normal in the following UL subframe. Once again, the superiority of KLD over L1D as a divergence metric is demonstrated by the larger peaks and lower baseline noise levels.

4.1.3. Test Signal C. The third signal used to test the detection capabilities of the divergence-based algorithm is shown in Figure 7. It depicts communication between a mobile terminal and base station using the WiBro standard. Since the recording is made at the MT, there is significantly more power in the UL subframes. Although unnoticeable in the time series, the UL subframe at 20 ms contains an additional command sequence that triggers the subsequent change in the transmit power observed at 25 ms. As a result, there are effectively two unusual events in the signal: the extra command and the subsequent change in power level. The parameters utilised for analysis of the signal are identical to those used in Section 4.1.2.

The divergence analysis plot in Figure 7 shows that both anomalies can be successfully detected using KLD and L1D. Since the width of a KLD peak corresponds to the temporal duration of the anomaly responsible, the first peak at 20 ms is very sharp as it is due to the extra command sequence in the UL subframe. Since the subsequent change in power at 25 ms affects the entire UL subframe, the second peak is much broader and spans the entire subframe.

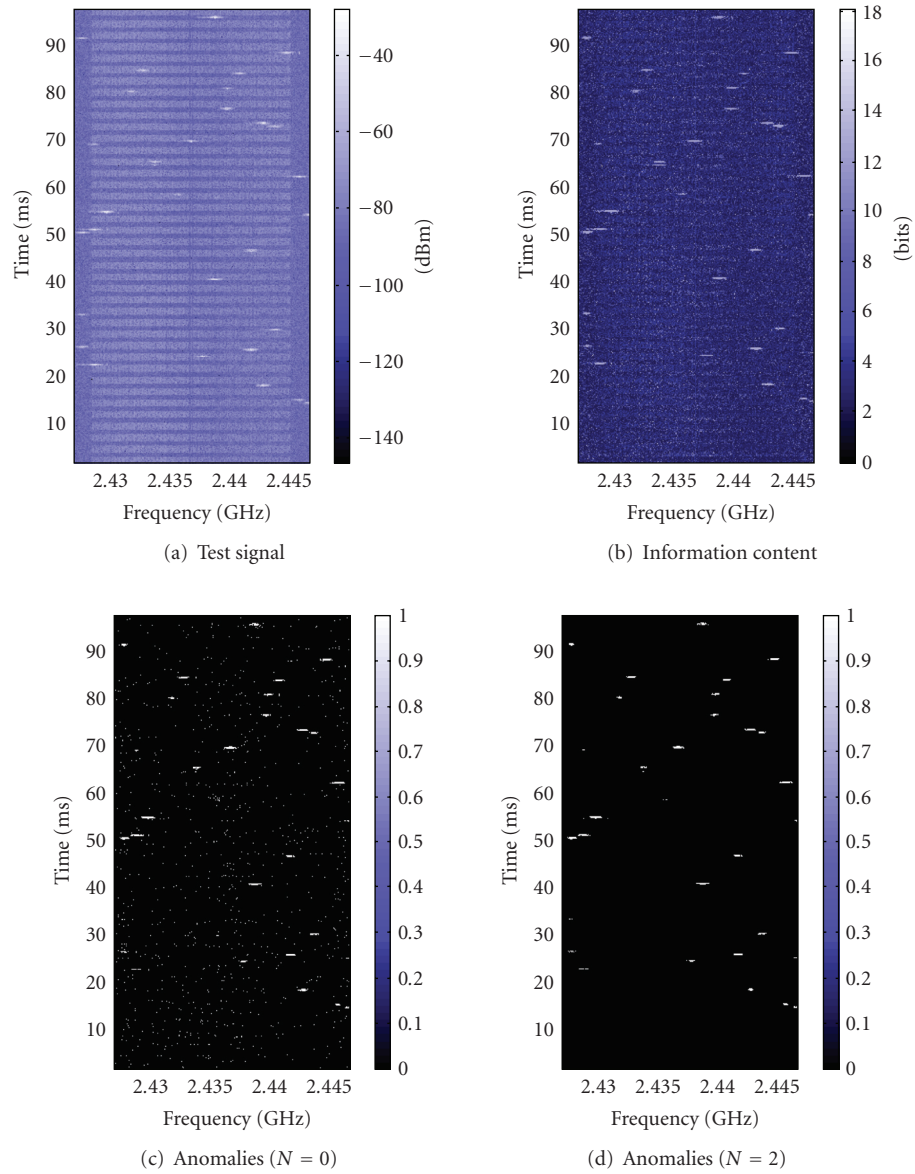


Figure 8: Analysis of a WLAN signal with interference from a Bluetooth device. (a) Spectrogram of the original signal. Data bursts from the Bluetooth device are clearly visible as high power, lightly shaded patches. (b) Information content of events (power density at any given time-frequency point) in the signal. (c) Anomalies detected (light patches) using a threshold of $1.25\sigma_{\mathrm{I}\langle 16\rangle}$ (7.15 bits) and a cluster length of 1. The result is noisy and there are a lot of false positives (appearing as singular, lightly shaded spots). (d) Anomalies detected when the threshold is left unchanged at $1.25\sigma_{\mathrm{I}\langle 16\rangle}$ and the cluster length is increased to 3. The outcome is now much cleaner with virtually zero false positives.

The plot also reveals that for the second anomalous event (power change), KLD is smaller than the associated L1D. This is not unexpected since it has been hypothesised in Section 3 that for differences that lead to an L1D of $2\ln 2$ or smaller, L1D can be larger than KLD. It is also the reason why L1D generally leads to larger baseline noise levels compared to KLD.

4.1.4. Test Signal D. The final test signal is used to evaluate the ICA algorithm. It is similar in nature to the signal shown in Figure 1(a). Spectrogram of the test signal is shown in Figure 8(a). It is a much longer signal with numerous interference events to provide a statistically significant sample size. The plot depicts a real WLAN signal with Bluetooth interference captured over the air-interface. The WLAN signal consists of a single frame that is repeated periodically by a vector signal generator. The characteristic frequency hopping pattern of the Bluetooth device marks the locations of the interference (anomalous) events.

The signal spectrogram is estimated from the time series using nonoverlapping Hamming windows that are $64\,\mu$s long. A 1024 point FFT (fast Fourier transform) is used to obtain a frequency resolution of approximately 20 kHz. The signal event under observation is the instantaneous power

density at any given time-frequency coordinate. The first 10 ms of the signal is assumed to be free from interference and is therefore used for training purposes. 16 equally spaced histogram bins divide the range between the maximum and minimum power densities observed in the training data. The $\sigma_{I\langle 16\rangle}$ for the training data is 5.72 bits.

Figure 8(b) shows the information content of events in the test signal. As expected, the anomalous events have a higher information content and they are highlighted while the regular underlying structure is suppressed. The plot also shows that there is a lot of noise (tiny spots of high information content) from individual low-probability signal events that are otherwise nominal. The reason for this behaviour has been outlined in Section 3.4.

Anomalies detected using a threshold of $1.25\sigma_{I\langle 16\rangle}$ and a cluster length of 1 (i.e., only the current event) are shown in Figure 8(c). It is immediately obvious from the large number of small, lightly shaded spots that there are a lot of false positives. Again, singular low-probability signal events are responsible since they can potentially have higher information content than actual anomalous events. Keeping I_{th} the same and increasing the cluster length to 3 yields the result shown in Figure 8(d). It reveals that a simple change in the cluster length is sufficient for reducing the number of false positives to virtually zero.

4.2. Performance Analysis. The analyses of test signals presented in Section 4.1 show that both algorithms perform well for the parameter combinations chosen. In order to investigate and quantify the impact of other parameter choices, it is necessary to define and utilise metrics that reflect performance.

For the divergence-based technique, the ratio between the anomaly detection peak and the maximum of the baseline noise level is a good indicator of performance since it is a reflection of the range over which a threshold can be applied. It can be seen from the results presented in Figures 5, 6, and 7 that KLD is an extremely effective discriminator for statistical changes in the observed data. Even with such *real test vectors captured over-the-air*, the KLD peaks produced by anomalous events are many orders of magnitude larger than baseline noise levels associated with nominal data. As a result, 100% probability of detection can be achieved over a wide range of KLD threshold values (the anomalous peak is approximately 140 times as large as the background noise level in Figure 6) while still guaranteeing a 0% probability of false positives—making such classical measures of detector performance inadequate for gauging the true extent of the algorithm's performance.

Another reason against the suitability of classical performance measures such as receiver operating characteristic (ROC) curves is the scarcity of available test data. Probability of detection and false positives are inherently statistical measures of performance that require a large sample size to produce meaningful results. Since the focus of this work is exclusively on practical applications of the proposed algorithm, the number of test vectors available is limited and each test signal (A, B and C) contains only 1 or 2 anomalous events. So instead of attempting to extract questionable probability measures from the limited data set, a measure of the difference between the height of the anomalous peak and the baseline noise level is utilised to quantify the observed performance.

When KLD is used as the measure of divergence, the KLD ratio (KLDR) is defined as

$$\mathrm{KLDR} = \frac{\mathrm{KLD}_{\mathrm{anom}}}{\mathrm{KLD}_{\mathrm{bg}}}, \tag{10}$$

where $\mathrm{KLD}_{\mathrm{anom}}$ is the maximum of the detection peak and $\mathrm{KLD}_{\mathrm{bg}}$ is the maximum of the background baseline noise level. KLDR is the metric that is used to quantify the algorithm's performance.

For the ICA-based algorithm, the circumstances are different. The test set (signal D) contains a sufficient amount of nominal and anomalous events to allow the use of more traditional performance metrics. Performance is measured in terms of the detector true positive rate (R_{tp}) and false discovery rate (R_{fd}). R_{tp} is defined as the ratio of the number of correctly detected anomalous events (Σ_{tp}) to the total number of anomalous events present (Σ_{ta}):

$$R_{\mathrm{tp}} = \frac{\Sigma_{\mathrm{tp}}}{\Sigma_{\mathrm{ta}}} = 1 - \frac{\Sigma_{\mathrm{md}}}{\Sigma_{\mathrm{ta}}}, \tag{11}$$

where Σ_{md} is the number of anomalous events that missed detection. R_{fd} is the ratio of false positives (Σ_{fp}) to the total number of anomalies detected (includes both Σ_{fp} and Σ_{tp}) [29]:

$$R_{\mathrm{fd}} = \frac{\Sigma_{\mathrm{fp}}}{\Sigma_{\mathrm{fp}} + \Sigma_{\mathrm{tp}}}. \tag{12}$$

R_{fd} is preferred over the more common false positive rate (R_{fp}) as it is more useful in this context. R_{fp} is defined as the ratio between Σ_{fp} and all nonanomalous events (Σ_{tn}) in the signal:

$$R_{\mathrm{fp}} = \frac{\Sigma_{\mathrm{fp}}}{\Sigma_{\mathrm{tn}}}. \tag{13}$$

It is also known as the false alarm rate. Since Σ_{tn} is a very large number, R_{fp} is close to zero for most parameter combinations and therefore does not adequately reflect the variations observed in detector performance.

To summarise, KLDR is used to evaluate the performance of the KLD-based algorithm while R_{tp} and R_{fd} are used to evaluate the ICA-based algorithm.

4.2.1. Sampling Rate. Continuous processes such as time-series must be sampled before the anomaly detection algorithms can be applied. The sampling frequency employed is crucial as it dictates the size of the input data sets, $|P_n|$, and therefore the memory utilisation of the KLD-based algorithm—as indicated in Table 2. For a given window length, a higher frequency implies that more data samples have to be stored and sorted to construct the histograms. If the frequency is too low, small scale signal features and anomalies may be lost. According to the Nyquist sampling

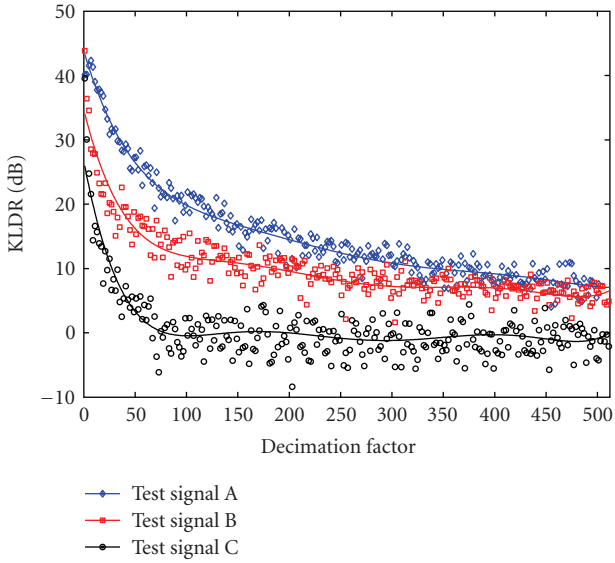


Figure 9: Performance of the KLD-based anomaly detection scheme under various data sampling rates. Decimation factor refers to the amount by which the input signal is undersampled relative to the signal bandwidth. A factor of unity corresponds to a sampling frequency equal to the signal bandwidth. For two of the test cases (A and B), a KLDR well above 10 dB can be maintained for a window length of 256 μs, histogram bin count of 32, and a decimation factor of 100. Signal C is unable to accommodate such high decimation rates.

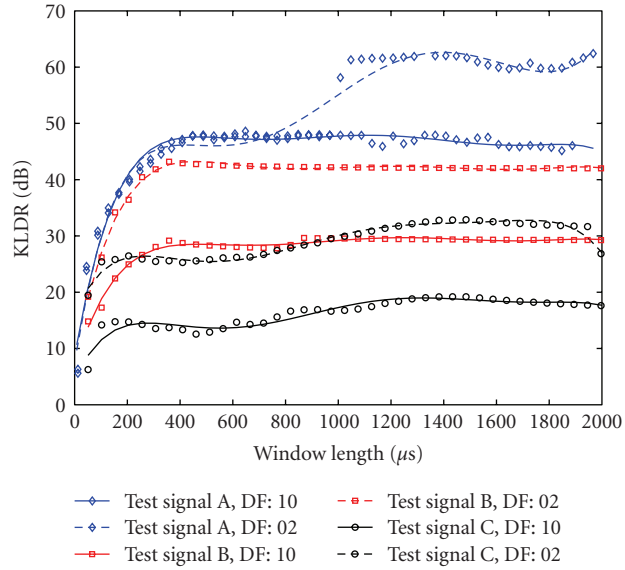


Figure 10: Performance of the KLD-based anomaly detection scheme under various window lengths and hence the input sample size $|P_n|$. Decimation factors (DF) of 2 and 10 are utilised and the number of histogram bins is 32. In all cases, an increase in the amount of data leads to a better performance. However, beyond a certain window length, the performance is no longer strongly affected.

criterion, a signal must be sampled with a frequency at least twice as large as its bandwidth to be reconstructible. For wideband signals this leads to a very high sampling frequency and hence a prohibitively large volume of data—heavily increasing the resource requirements of the proposed scheme. Since neither of the proposed algorithms require the time-series to be reconstructible, a far lower sampling frequency can be used instead. Figure 9 shows how the performance of the KLD-based anomaly detection scheme is affected by undersampling of the input time series. The window length utilised is 256 μs and the histograms used to construct the PMFs are 32 bins wide. The amount by which the input time-series is undersampled relative to the bandwidth is defined as the decimation factor. Therefore, a factor of unity implies that the signal is sampled at the same frequency as the signal bandwidth.

The results indicate that decimation factors as large as 500 can be successfully employed depending on the type and duration of the anomaly present. For test signals A and B, a KLDR of more than 10 dB can be maintained even with a decimation factor of 100. This is an important result as it indicates that satisfactory performance levels can be maintained with little input data and hence memory-limited implementations of the algorithm. At high decimation factors, performance is poor for test signal C. This is because the first anomaly (extra command sequence) is temporally brief and is likely to be lost when the signal is heavily undersampled. As for the second anomaly in the signal, the change in power is simply not large enough to produce a significant increase in the divergence.

Results for the ICA-based method are not shown since the input sampling rate has no bearing on the complexity or memory requirements of the algorithm—as shown in Table 4. The only requirement then on the sampling rate is that it must be fast enough to capture events that are suspected of being anomalous.

4.2.2. Window Length. The window size, and hence the input data set size $|P_n|$, is another parameter that is relevant for the KLD-based algorithm but not the ICA-based algorithm. The effect of the PMF estimation window size on the performance of the algorithm shown in Figure 10. The number of histogram bins utilised is 32. Results are shown for undersampling factors of 2 and 10. At lower decimation factors, more data is available and the KLDR improves uniformly across all window sizes for signals A and B. At smaller window sizes, performance for signal A is unaffected by the choice of the decimation factor due to the relatively long duration of the anomaly. This is because even at a decimation factor of 10, a sufficient number of anomalous samples are represented in the PMF.

As anticipated, the performance is poor at small window sizes where the amount of data available is insufficient to adequately model the underlying PMFs. Increasing the window length leads to an improvement of the performance. However, for signals B and C, the gains become marginal for windows larger than approximately 400 μs. The transition shown by signal A at a window length of 1 ms for a decimation factor of 2 is due to a sudden reduction in the KLD noise at the frame edges (as seen in the zoomed-in

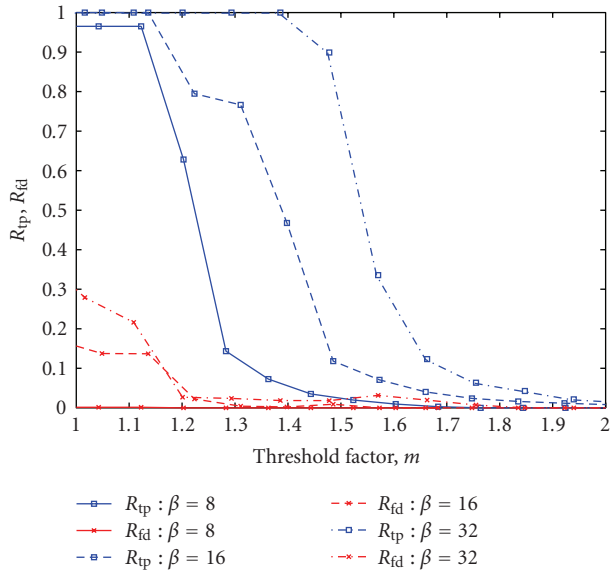


FIGURE 11: Effect of information threshold (9) on detector R_{tp} and R_{fd} for the ICA algorithm. The cluster length utilised is 3. Histogram bin sizes of 8, 16, and 32 are used for comparison. R_{tp} for each β shows a sharp decrease when I_{th} (i.e., $m \cdot \sigma_{I\langle\beta\rangle}$) is increased above a certain limit. Majority of the anomalous events have an information content less than this and miss detection. $m = 1.35$ and $\beta = 32$ yields the best performance.

segment of Figure 5) while the detection peak remains at approximately the same level. It is no coincidence that the duration of the interframe spacing for the signal is also 1 ms. It is a signal feature that is detected by the algorithm alongside the actual anomalous events. When the window size is increased beyond this feature size, it can no longer be resolved effectively by the detection algorithm and lead to a decrease in the noise level seen at the frame edges.

The initial KLDR improvements with increasing window size are due to improvements in the PMF estimates which in turn lead to a reduction in the baseline KLD levels. At larger window sizes, the anomalous samples represent smaller fractions of the data and hence contribute less to the shape of the estimated PMF—resulting in a decrease of the KLD due to the anomaly. As the background levels are also reduced by an increase in the data size, the overall KLD ratio (i.e., the KLDR) remains relatively constant.

4.2.3. Information Threshold. As stated previously in Section 3, the discrimination threshold (9) is an important aspect of any detector. The impact of I_{th} on R_{tp} and R_{fd} of the ICA-based algorithm is investigated using a cluster length of 3 and histogram bin sizes of 8, 16, and 32. The result of the analysis is shown in Figure 11.

The plot shows that there is a hard I_{th} boundary for each β after which R_{tp} drops rapidly. This implies that the majority of the anomalous events share similar characteristics and convey information equivalent to that boundary. When I_{th} is increased further through the use of a larger threshold factor m, R_{tp} approaches zero due to an ever increasing number of missed detections.

At low-information content thresholds, R_{fd} is also high—specially for high values of β. As explained earlier in Section 3.1, a higher resolution makes the detector more susceptible to noise, leading to an increase in the number of false positives and hence the R_{fd}.

The impact of D_{th} on the performance of the KLD-based method has not been investigated and therefore cannot be shown. The reasons for this are as follows.

(i) The number of anomalous events available is insufficient to investigate statistical trends.

(ii) Detection is often guaranteed for a wide range of thresholds due to the large KLDR (greater than 30 dB) values that are observed.

It is the second reason that generally makes it straightforward to choose a suitable D_{th}.

A mathematical treatment of the impact of D_{th} and I_{th} on detector performance is beyond the scope of this article. Such a framework requires well-defined theoretical models of the data distribution which are difficult to obtain for real data vectors. Equations derived using the simplifying assumption that the distributions belong to a well-known class such as Gaussian would be of little use in context of the test signals used in this paper. Since the signals do not conform to any standard probability density function, it is out of necessity that the thresholds are determined empirically.

4.2.4. Histogram Resolution. Histogram bin resolution, represented by the parameter β, is of relevance to both of the proposed algorithms. Figure 12 shows how performance of the KLD-based algorithm is affected by the choice of the number of histogram bins used to classify the input data and estimate the PMFs. The window length is set at 256 μs and results are shown for decimation factors of 2 and 10. Once again, the smaller decimation factor provides uniformly improved performance over the entire range of β values. The only exception is signal A where the performance for smaller β values appears to be independent of the decimation factor used. The reasons for this is the relatively long duration of the anomaly—as explained previously in Section 4.2.2.

The only trend common to all three signals is that the performance changes little with increasing bin numbers, with signal C showing an optimum in the vicinity of $\beta = 55$. This indicates that the behaviour observed is specific to the type of anomaly present in a signal. While the number of bins utilised does not appear to have a significant impact on the performance of the scheme for a fixed amount of data, the decrease observed is due to noisier PMF estimates that are obtained for larger values of β. Noisy PMFs lead to larger background KLD values and hence a reduced KLDR.

Figure 13 shows how the R_{tp} and R_{fd} vary for the test signal (Figure 8(a)) with the number of histogram bins utilised. The cluster length utilised is 3 and I_{th} of $1.2\sigma_{I\langle\beta\rangle}$, $1.4\sigma_{I\langle\beta\rangle}$, and $1.6\sigma_{I\langle\beta\rangle}$ are used for comparison. The plot reveals that when $\beta = 4$, R_{tp} is zero and R_{fd} is unity for all thresholds tested. This is because the sensitivity is very low and no anomalies can be detected ($R_{tp} = 0$). Events exceeding the threshold are low-probability signal events and

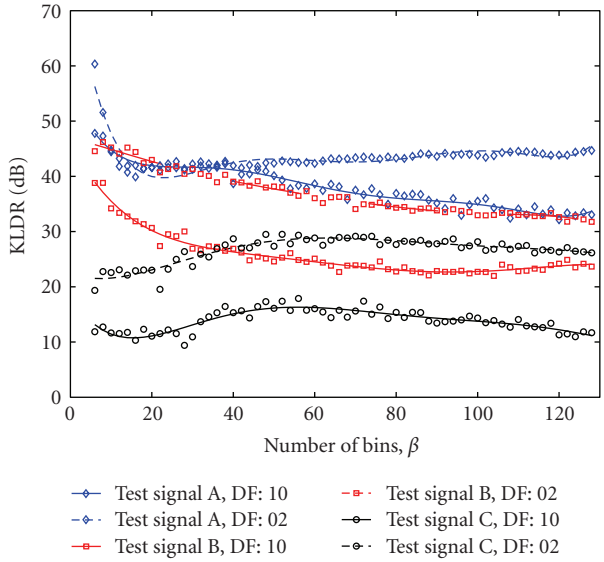


FIGURE 12: Performance of the KLD-based anomaly detection scheme under various histogram bin counts. The window size is set at 256 μs and the decimation factors used are 2 and 10. Generally, a larger number of bins lead to poorer performance due to increased noise in the estimates. However, the rate of change is small and therefore the drop in performance is insignificant over a wide range of bin resolutions.

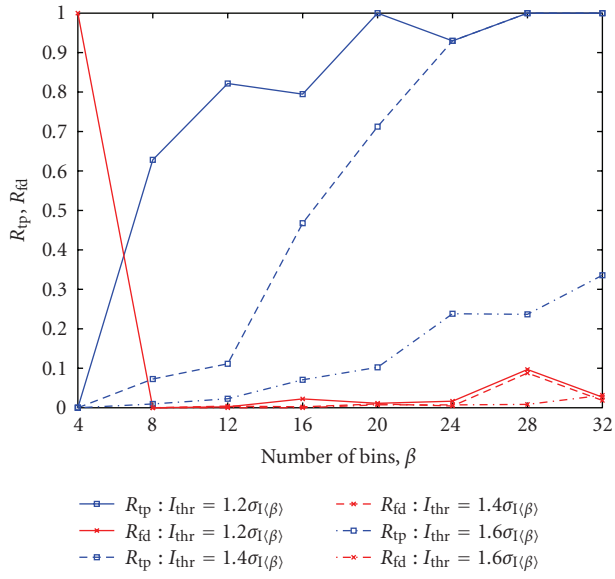


FIGURE 13: Effect of histogram resolution on detector R_{tp} and R_{fd}. The cluster length is 3 and thresholds are $1.2\sigma_{I\langle\beta\rangle}$, $1.4\sigma_{I\langle\beta\rangle}$, and $1.6\sigma_{I\langle\beta\rangle}$. R_{tp} improves with resolution while R_{fd} deteriorates. $\beta = 20$ with a threshold of $1.2\sigma_{I\langle 20\rangle}$ yields the best performance.

hence are all false positives ($R_{fd} = 1$). As β is doubled to 8, the resolution improves and there is a corresponding increase in the R_{tp}. The R_{fd} also drops to a negligibly small value. As β is increased further, the R_{tp} increases due to better detector resolution. The R_{tp} improvements come at a cost, the detector is more susceptible to noise at higher

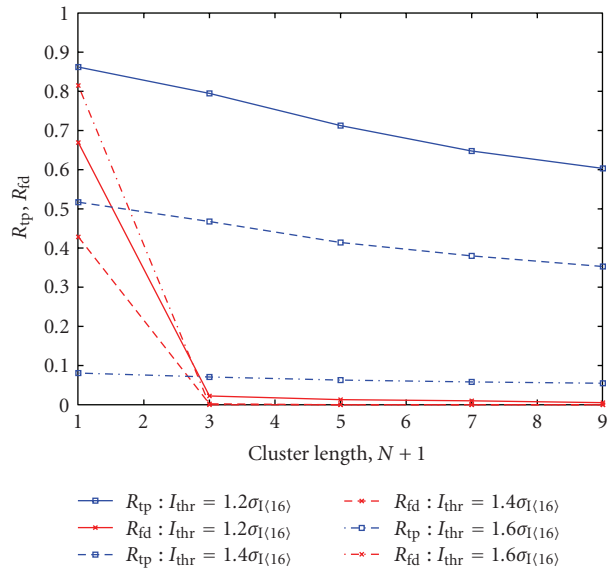


FIGURE 14: Impact of cluster length on detectors R_{tp} and R_{fd}. The number of histogram bins used is 16 and the thresholds utilised are $1.2\sigma_{I\langle 16\rangle}$, $1.4\sigma_{I\langle 16\rangle}$, and $1.6\sigma_{I\langle 16\rangle}$. Even the smallest cluster length ($N = 2$) is shown to provide a significant improvement in the R_{fd}.

resolutions. This is evident from the gradual increase in the R_{fd}.

Comparison between the three detection thresholds reveals that a higher R_{tp} is achieved with a lower threshold. Unfortunately, this also leads to a higher R_{fd}. This behaviour is in accordance with the explanation provided in Section 4.2.3.

4.2.5. Cluster Size. In order to investigate the impact of the information cluster length, N, on the ICA-based detector, β is set at 16 and the analysis is performed for I_{th} of $1.2\sigma_{I\langle 16\rangle}$, $1.4\sigma_{I\langle 16\rangle}$, and $1.6\sigma_{I\langle 16\rangle}$ on the test signal shown in Figure 8. The result of the analysis is shown in Figure 14.

The significance of clustered anomaly detection is immediately obvious. With $N = 0$, when clustering is not performed, there are an overwhelming number of false positives. This is indicated by the high R_{fd}. As soon as clustering is applied by setting $N = 2$, a dramatic drop in the R_{fd} is observed—showing that even minimal anomaly clustering is sufficient to yield a massive improvement in detector performance. By lowering the R_{fd}, clustering also allows a lower I_{th} to be used to achieve a higher R_{tp}.

The impact of anomaly clustering on R_{tp} for a given I_{th} is relatively low. As cluster size is increased, a gradual decrease is observed in the R_{tp}. This is expected since larger cluster sizes lead to missed detections around the edges of the interference patterns. The plot also shows that higher thresholds lead to lower R_{tp} for a given cluster size. This is also expected since a higher information content threshold leads to a higher number of missed detections.

From the analysis performed on the test signals, it is clear that it is challenging to determine a set of parameters that are inherently optimal for the anomaly detection algorithms

proposed. This is due to the fact that the optimal parameter set depends on a number of problems-specific factors such as the duration of the anomaly and the dynamic range of the signal. It may be possible to develop adaptive variants of the algorithms that automatically find the best parameter combinations subject to some performance criterion but that is beyond the scope of this paper.

The ICA-based algorithm is particularly sensitive to the parameters utilised. Generally, it is seen that parameter values that increase the R_{tp} (good) often also lead to an increase in the R_{fd} (bad) and vice versa. Trade-offs must therefore be made to meet the required detector performance characteristic (low R_{fd}, moderate R_{tp} or high R_{tp}, moderate R_{fd}). A moderate number of bins ($\beta = 20$), small cluster size ($N = 2$), and a threshold of $1.2\sigma_{I\langle 20\rangle}$ bits ($m = 1.2$) provide a good balance between R_{tp} (1.0) and R_{fd} (0.01) for this particular test vector (signal D).

The KLD-based algorithm on the other hand is much more robust with respect to the parameter combinations utilised. The results clearly show that performance better than 30 dB of KLDR can be easily obtained with reasonable choice of parameter values such as a window length of 256μs, $\beta = 32$ and decimation factor of 2.

5. Hardware Platform

The FP-DKLD version of the detection algorithm presented in Section 3.3 has been implemented on a Xilinx Virtex-4 ML402 SX XtremeDSP Evaluation Platform to serve as a proof of concept and allow the testing of signals in real time. To facilitate and accelerate code development, Xilinx SystemGenerator 10.1 is used in conjunction with MATLAB R2007a for the primary design flow. The implemented design runs at a clock speed of 80 MHz and is capable of processing input with a 10 MHz sample rate. The hardware chain used to test and validate the FP-DKLD implementation is shown in Figure 15.

The Agilent E4438C ESG signal generator simultaneously provides analogue and digital versions of the signal under test. The digital data stream is connected to the FPGA platform via the Agilent N5102A Digital Signal Interface Module (DSIM) while the analogue signal is connected to a oscilloscope for display. The DSIM conditions the data (word size, bit alignment, clock relationship settings) and provides a synchronous clock signal that is used to drive the FPGA core. The trigger output from the FPGA platform is also connected to the oscilloscope via a digital probe so that it can be directly compared against the signal under test.

A pair of Wireless Broadband (WiBro) signals known to contain a number of different anomalous data segments are used to test the hardware platform. The design is configured with $|P_n| = 4096$ (320 μs) and $\beta = 8$. It is not necessary to down sample the input data stream since the implemented design is capable of processing the input at its original rate.

The DSIM module provides the samples to the FPGA as 12-bit words in 2s complement format. The sample and DSIM clocks are set at 10 MHz and 40 MHz respectively-providing 4 clock cycles per input sample (CCPS). Although

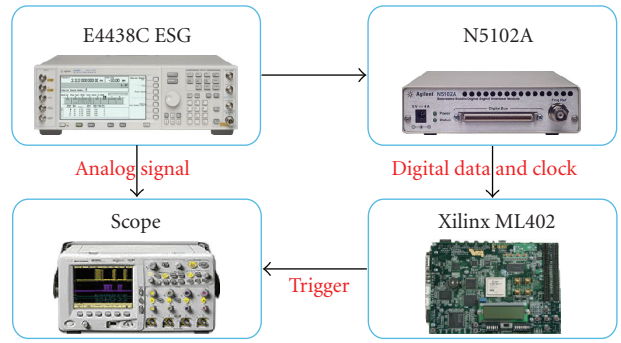


Figure 15: Block diagram of the hardware test-bed. The E4438C ESG signal generator produces the signal under test in both analogue and digital formats. The digital signal is passed to the FPGA platform via the N5102A digital signal interface module while the analogue signal is fed into the oscilloscope. The trigger signal from the FPGA core is also connected to the oscilloscope via a digital probe for comparison.

the design requires 8 CCPS, the DSIM is only capable of providing a maximum of 4 CCPS. To obtain the required 8 CCPS, the clock signal is doubled on the FPGA using an on-chip digital clock manager (DCM) module. Use of a DCM also has added benefit of providing clock buffering and deskewing.

5.1. Test Signal I. Figure 16 shows the result of analysing the first WiBro signal using the FP-DKLD implementation of the algorithm. The signal analysed is identical to that shown in Figure 6 and analysed in Section 4.1.2. It is clear from the oscilloscope trace that one of the UL frames is longer than the others and hence is anomalous. With $D_{th} = 0.0313$, the FPGA implementation of the algorithm clearly succeeds in detecting the signal anomaly. The first trigger event obtained (A) coincides exactly with the anomalous segment of the unusual UL frame. A second trigger event (B) is observed when the UL frame structure subsequently returns to normal and the anomalous segment is no longer present.

5.2. Test Signal II. The second WiBro signal tested is shown in Figure 17. It is identical to that shown in Figure 7 and analysed in Section 4.1.3 with the exception of an additional change in the timing structure. Analysing the signal with $D_{th} = 0.0625$ is seen to produce five trigger events—corresponding to the three anomalous conditions known to be present in the signal.

Trigger events A and B are due to a momentary disruption in the natural frame period of the signal. The first event marks the position where the UL frame should have been but is not while the second event marks the opposite: finding a UL frame where there should be none.

Events C and D are caused by a very brief command sequence at the beginning of the fifth UL frame that causes the power control loop to be initiated—which is then responsible for event E. Although invisible to the naked eye, the algorithm succeeds in locating the anomalous command sequence as clearly demonstrated by trigger event

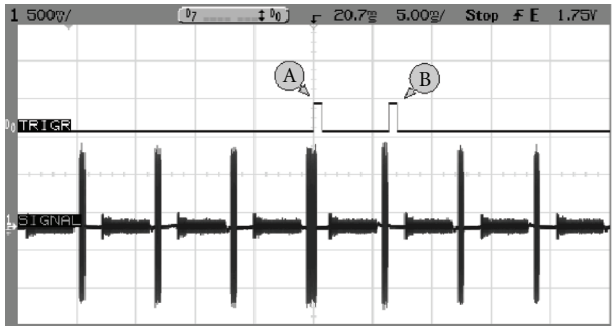


Figure 16: Oscilloscope trace of a WiBro signal with a single anomalous frame and the associated trigger events. The KLD threshold is 0.0313. Trigger event A marks the start of the unusual segment of the anomalous frame. A second trigger event, B, is also obtained in the subsequent frame due to the disappearance of the anomalous feature.

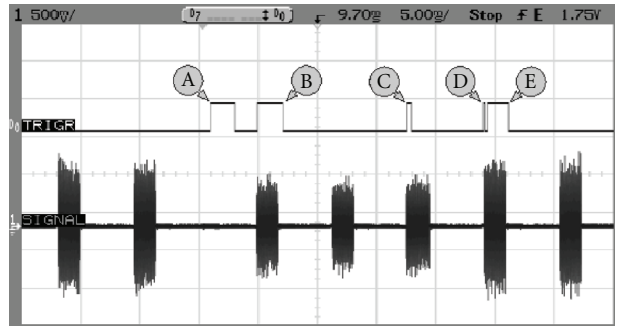


Figure 17: Oscilloscope trace of a WiBro signal with multiple anomalous events and the associated trigger events. The KLD threshold is 0.0625. Triggers A and B are caused by a momentary change in the signal period. Trigger C flags the presence of a very brief command sequence that leads to a signal power level change. Trigger D marks the position in the subsequent frame where the power change inducing command was previously present and finally, trigger E corresponds to a sudden change in the overall signal power level.

C. Since that command sequence is no longer present in the subsequent UL frame, its disappearance is marked by trigger event D.

Once initiated, the power control loop causes a sudden increase in the signal power level. This behaviour can be considered to be anomalous and is flagged by trigger event E. There are no other events associated with the change in power level as the signal power is seen to remain high beyond this point.

6. Summary and Conclusions

Two complementary anomaly detection algorithms utilising information theoretic measures have been presented. Both algorithms are simple to implement and require little a priori information regarding the signal under test. Demodulation of the signal is also not required since the algorithms are capable of processing the baseband signal envelope itself in real time. The information content analysis based method is capable of detecting singular anomalous events while the Kullback-Leibler divergence based method is also able to detect otherwise nominal events that are anomalous purely due to context (e.g., misaligned signal frames). In order to provide this context aware detection of anomalies, the KLD-based algorithm requires the input signal to be periodic.

Analyses of a number of test signals captured over the air show that the KLD-based scheme is successful at detecting all anomalies known to be present. Extensive tests using a software implementation of the algorithm demonstrate that it is robust with respect to parameter choices since satisfactory performance can be maintained with reasonable parameter values even when the input is severely undersampled. With PMF estimation window sizes of 256 μs, 32 histogram bins and factor of 10 undersampling, KLDR of 25 dB or better can be achieved depending on the anomaly present.

Although the primary purpose of the KLD-based algorithm is to act as an anomaly detector, it can also be used to detect frame boundaries in a signal. The modification required is trivial: eliminate the spacing that normally separates the two PMF estimation windows. Since frame boundary detection is expected to reveal the underlying cyclic structure of a periodic signal, it may be used as a precursor to the actual anomaly detection algorithm to automatically learn the period of the signal—thus eliminating the need for any a priori information regarding a test vector.

A variation on the anomaly clustering technique presented in context of the ICA-based algorithm may also be applied to the KLD-based algorithm to further improve detection of anomalous events. Anomalies generally lead to KLD peaks that increase monotonically until some maximum divergence is reached. It may be possible to exploit this observation to improve detection under low SNR conditions by restricting detection to signal segments that lead to monotonically increasing KLD values that are also above some predefined KLD threshold.

Both boundary detection and monotonic sequence detection are techniques that add significantly to the KLD-based anomaly detection algorithm. Therefore, they will be the primary focus of work done in the future on this subject.

In addition to the MATLAB based software, the algorithm has been implemented on a Xilinx Virtex4 FPGA based hardware platform for evaluation under real world physical conditions. The design is highly efficient and capable of processing 10 MHz input signals without requiring any undersampling. Successful tests with a set of WiBro signals indicate that the algorithm is indeed capable of processing high speed test vectors in real time.

Unlike the KLD-based method, the algorithm utilising ICA for anomaly detection does not require the input signal to be periodic. The only piece of information that is needed in advance is the set of reference event probabilities. A training data set known to be clean can be used to obtain the reference probabilities prior to analysis. The complexity and memory requirements of the algorithm are also very low.

It is clear from tests carried out using a software implementation of the algorithm that performance of the

system is strongly affected by the choice of parameters such as histogram resolution, threshold and cluster size. Impact of these parameters on the detector performance by means of the true positive rate and false discovery rate has been analysed and guidelines for appropriate values have been provided. It is shown that a true positive rate of 100% and false discovery rate of 1%—guaranteeing zero missed detections with very few false positives—is possible for the signal tested with a suitable set of parameter choices.

The ICA-based algorithm presented in this paper utilises a histogram with infinite memory, that is, it maintains a record of all samples analysed. Clearly this implies that the information content of anomalous events such as an interference drops over time if they happen with sufficient frequency. If such behaviour is undesirable, it is necessary to implement a windowed histogram. Along with tests against other types of anomalous signals, it will be the focus of further research on the ICA algorithm.

Acknowledgments

The authors would like to thank Roy Macnaughton and Peter Cain for the provision of test data. They greatly acknowledge the financial support of this project from Agilent Technologies and the University Relations Program. Harald Haas acknowledges the Scottish Funding Council support of his position within the Edinburgh Research Partnership in Engineering and Mathematics between the University of Edinburgh and Heriot Watt University.

References

[1] H. Shimazaki and S. Shinomoto, "A method for selecting the bin size of a time histogram," *Neural Computation*, vol. 19, no. 6, pp. 1503–1527, 2007.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 849–877, 2009.

[4] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 772–776, IEEE, Pacific Grove, Calif, USA, November 2004.

[5] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," in *Proceedings of the IEEE Military Communications Conference (MILCOM '06)*, pp. 1–7, Washington, DC, USA, October 2006.

[6] M. Öner and F. Jondral, "Air interface recognition for a software radio system exploiting cyclostationarity," in *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, vol. 3, pp. 1947–1951, Barcelona, Spain, 2004.

[7] C. Krügel, T. Toth, and E. Kirda, "Service specific anomaly detection for network intrusion detection," in *Proceedings of the ACM Symposium on Applied Computing*, pp. 201–208, ACM, Madrid, Spain, March 2002.

[8] J. Lin, E. Keogh, A. Fu, and H. Van Herle, "Approximations to magic: finding unusual medical time series," in *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*, pp. 329–334, IEEE, Dublin, Ireland, June 2005.

[9] R. J. Bolton and D. J. Hand, "Statistical fraud detection: a review," *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002.

[10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, edited by D. L. Schilling, John Wiley & Sons, New York, NY, USA, 1st edition, 1991.

[12] J. G. Proakis, *Digital Communications*, McGraw-Hill Higher Education, New York, NY, USA, 4th edition, 2000.

[13] J. Robertson, E. W. Tallman, and C. H. Whiteman, "Forecasting using relative entropy," FRB of Atlanta Working Paper No. 2002-22, November 2002, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=355460.

[14] H. Nakahara and S.-I. Amari, "Information-geometric measure for neural spikes," *Neural Computation*, vol. 14, no. 10, pp. 2269–2316, 2002.

[15] D. D. Falconer, F. Adachi, and B. Gudmundson, "Time division multiple access methods for wireless personal communications," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 50–56, 1995.

[16] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

[17] M. Markou and S. Singh, "Novelty detection: a review—part 1: statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481–2497, 2003.

[18] M. Markou and S. Singh, "Novelty detection: a review—part 2: neural network based approaches," *Signal Processing*, vol. 83, no. 12, pp. 2499–2521, 2003.

[19] A. Lazarevic, L. Ertöz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection," in *Proceedings of the SIAM International Conference on Data Mining*, pp. 25–36, San Francisco, Calif, USA, May 2003.

[20] Z. A. Bakar, R. Mohemad, A. Ahmad, and M. M. Deris, "A comparative study for outlier detection techniques in data mining," in *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, pp. 1–6, IEEE, Bangkok, Thailand, June 2006.

[21] S. Basu and M. Meckesheimer, "Automatic outlier detection for time series: an application to sensor data," *Knowledge and Information Systems*, vol. 11, no. 2, pp. 137–154, 2007.

[22] M. J. Desforges, P. J. Jacob, and J. E. Cooper, "Applications of probability density estimation to the detection of abnormal conditions in engineering," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 212, no. 8, pp. 687–703, 1998.

[23] D.-Y. Yeung and C. Chow, "Parzen-window network intrusion detectors," in *Proceedings of the International Conference on Pattern Recognition*, vol. 4, pp. 385–388, IEEE, Quebec City, Canada, 2002.

[24] W. Lee and D. Xiang, "Information-theoretic measures for anomaly detection," in *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 130–143, IEEE, Los Alamitos, Calif, USA, May 2001.

[25] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, no. 4, pp. 349–369, 1989.

[26] R. Krichevsky and V. Trofimov, "The performance of universal encoding," *IEEE Transactions on Information Theory*, vol. 27, no. 2, pp. 199–207, 1981.

[27] D. H. Johnson, C. M. Gruner, K. Baggerly, and C. Seshagiri, "Information-theoretic analysis of neural coding," *Journal of Computational Neuroscience*, vol. 10, no. 1, pp. 47–69, 2001.

[28] H. Haas and S. McLaughlin, Eds., *Next Generation Mobile Access Technologies: Implementing TDD*, Cambridge University Press, Cambridge, UK, 2008.

[29] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B*, vol. 57, no. 1, pp. 289–300, 1995.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2010, Article ID 194306, 6 pages
doi:10.1155/2010/194306

Research Article

Frequency-Based Optimization Design for Fractional Delay FIR Filters with Software-Defined Radio Applications

Javier Díaz-Carmona,[1] Gordana Jovanovic-Dolecek,[2] and Agustín Ramírez-Agundis[1]

[1] *Electronics Department, Technological Institute of Celaya, Celaya, GTO 38010, Mexico*
[2] *Electronics Department, National Institute of Astrophysics Optics and Electronics, Tonantzintla, PUE 72840, Mexico*

Correspondence should be addressed to Javier Díaz-Carmona, jdiaz@itc.mx

Received 8 November 2009; Accepted 1 January 2010

Academic Editor: Massimiliano Laddomada

A frequency-designed fractional delay FIR structure, which is suitable for software radio applications, is presented. The design method is based on frequency optimization of a combination of modified Farrow and mutirate structures. As a result the optimization frequency range is made only in half of desired total bandwidth. According to the obtained results the proposed fractional delay structure allows online desired fractional delay update, with a high fractional delay value resolution.

1. Introduction

Software-defined radio represents a major change in the design paradigm for radios, [1], in which most of the functionality is made through programmable signal processing devices, giving the radio the ability to change its operating parameters to accommodate new features and capabilities [2].

A software-defined radio platform is designed to make mobile systems more flexible with respect to the bandwidth requirements of different mobile standards. This flexibility is achieved by performing channel selection in the digital domain through sample rate conversion (SRC) with programmable digital filters. Fractional Delay (FD) filters are key components used to perform nonrational SRC [3–5].

Additionally, Software radio systems employ direct conversion receivers with asynchronous sampling such that the actual sampling instants are not synchronized with the incoming symbols. In order to evaluate the received symbols a digital symbol synchronization is implemented through FD filter structures [6, 7].

One of the key requirements for FD in software radio applications is to have the flexibility to change among different communication protocols and to be able to perform fractional delay value update on line, known as variable fractional delay (VFD) filters [8]. Other important FD characteristics are wide bandwidth, high fractional delay resolution value, and a small number of arithmetic operations per output sample.

There are several FD design methods [9]; among them the use of a polynomial approach allows online desired fractional delay value update using a Farrow structure [10] or a modified Farrow structure [11, 12]. Both structures are composed of $L + 1$ parallel FIR filters $C_l(z)$, each one with length N, where L is the chosen polynomial order, as it is shown in Figure 1. In a modified Farrow structure $\gamma = 2\alpha - 1$, where α is the required fractional delay value, $0 < \alpha < 1$, and $C_l(z)$ are linear phase filters (symmetrical coefficient values). In the original Farrow structure $\gamma = \alpha$ and parallel filters are not symmetrical.

There are two main FD design approaches based on the polynomial approach. The first one is completely time domain design based on either Lagrange interpolation [13] or B spline functions [14, 15]. The implementation of this design approach is made through original Farrow structure having as main advantage that filter coefficients are obtained by closed form expressions. The disadvantage is the small flexibility available for this approach to meet FD filter

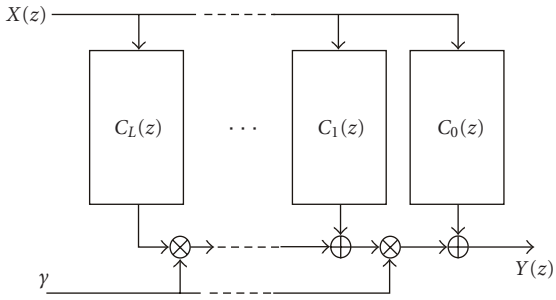


Figure 1: Farrow structure.

frequency domain specifications. This is because there is only one design parameter, that is, the polynomial order L. Most of the recently reported FD design methods belong to this time domain approach such as [16, 17].

The second FD design approach is made in frequency domain using optimization techniques for coefficients computing. The main advantage of this design approach is an improved control on frequency specifications. This is because three design parameters are available: polynomial order L, filter length N, and desired frequency passband ω_p. The disadvantage of this approach is the need of an optimization method for filter coefficient computing. Several design methods have been proposed such as, for example, [18], where the FD filter is implemented in a modified Farrow structure and a Taylor approximation is achieved. Similarly in [19–21] the implementation is made using an original Farrow structure and a weighted least square optimization is accomplished.

The use of frequency domain design methods for FD filters with a wide bandwidth requires that an optimization method be applied over a large frequency range. On the other hand large filter length N and polynomial order L are obtained when high fractional delay resolution is required. Hence the design process requires high coefficients computing time and a high number of arithmetic operations per output sample in the resulting FD filter implementation.

This paper describes the use of a multirate structure in a frequency design approach in order to reduce the optimization workload in coefficients computing for FD filters with a wide bandwidth, high fractional delay resolution, and online fractional delay value update capability. In this way a flexible frequency design method with a reduced optimization workload as well as a resulting structure with a reduced number of arithmetic operations per output sample is obtained.

The used frequency design method is the modified Farrow structure [18], where each parallel filter $C_l(z)$ is designed as a minimum least square approximation of an l order differentiator. In the same way it is possible to extend such proposal through other optimization frequency design methods.

Section 2 describes in a general way the frequency design method basis. The multirate structure is given in Section 3. In Section 4 the proposed design method is shown, which is illustrated through one design example. Conclusions are presented in Section 5.

2. Frequency Design

The frequency design method in [18] is based on the following properties of the parallel digital filters $C_l(z)$.

(1) FIR filters $C_l(z)$, $0 \le l \le L$, in original Farrow structure are an L order Taylor approximation to the continuous-time interpolated input signal.

(2) In the modified Farrow structure the FIR filters $C_l(z)$ are linear phase type II for l even and type IV for l odd.

Each filter $C_l(z)$ approximates in magnitude the function $K_l\omega^l$, where K_l is a constant. The ideal frequency response of an l order differentiator is $(j\omega)^l$; hence the ideal response of each $C_l(z)$ filter in the Farrow structure is an l order differentiator.

In the same way it is possible to approximate the input signal through Taylor series in a modified Farrow structure. The l order differential approximation to the continuous-time interpolated input signal is done through the branch filter $C_l(z)$, with a frequency response given as

$$\hat{C}_l\left(e^{j\omega}\right) = e^{-j\omega((N-1)/2)}\frac{(-j\omega)^l}{2^l l!}. \tag{1}$$

The input design parameters are the filter length N, the polynomial order L, and the desired passband frequency ω_p.

The N coefficients of the $L+1$ $C_l(z)$ FIR filters are computed in such a way that the following error function is minimized in a least square sense in the frequency range $[0, \omega_p]$:

$$e_l(\omega) = \left[\sum_{n=0}^{(N/2)-1} C_l\left(\frac{N}{2}-1-n\right)\gamma(l,n,\omega) - D(l,\omega)\right], \tag{2}$$

where

$$\begin{aligned} D(l,\omega) &= \frac{(-\omega)^l}{2^l l!}, \\ \gamma(l,n,\omega) &= 2\cos\left[\left(n+\frac{1}{2}\right)\omega\right], l \quad \text{even}, \\ \gamma(l,n,\omega) &= 2\sin\left[\left(n+\frac{1}{2}\right)\omega\right], l \quad \text{odd}. \end{aligned} \tag{3}$$

Hence the objective function is given as

$$E_1 = \int_0^{\omega_p}\left[\sum_{n=0}^{(N/2)-1} C_l\left(\frac{N}{2}-1-n\right)\gamma(l,n,\omega) - D(l,\omega)\right]^2 d\omega. \tag{4}$$

From this equation it can be observed that the design of a wide bandwidth FD filter requires an extensive computing workload. For high fractional delay resolution FD filters high precise differentiator approximations are required; this

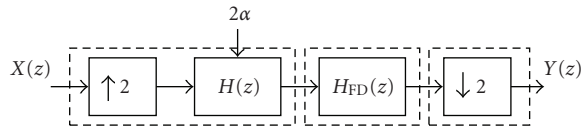


Figure 2: Multirate structure for FD filter.

implies high branch filters length, N, and high polynomial order, L. Hence an FD filter structure with high number of arithmetic operations per output sample is obtained.

The arithmetic complexity of the resulting implemented structure is an important factor to be considered. The comparative parameters are the following:

(1) number of multipliers per output sample (MPS),

(2) number of additions per output sample (APS).

In the modified Farrow structure the MPS_1 and APS_1 are given as

$$\begin{aligned} \text{MPS}_1 &= \left(\frac{N}{2}\right)(L+1)+L, \\ \text{APS}_1 &= (N-1)(L+1)+L+1. \end{aligned} \tag{5}$$

3. Multirate Structure

The multirate structure in [22] is proposed for designing FD filters in time domain. The input signal bandwidth is reduced by incrementing the sampling frequency. In this way Lagrange interpolation is used in filter coefficients computing for an FD filter with a wide bandwidth.

The multirate structure shown in Figure 2 is composed of three sections. The first one is an upsampler and a half band image suppressor filter $H(z)$ for incrementing twice the input sampling frequency. The second section is the FD filter $H_{\text{FD}}(z)$, which is designed in time domain through Lagrange interpolation [11]. Since the signal processing frequency of filter $H_{\text{FD}}(z)$ is two times more than the input sampling frequency, such filter can be designed to meet only half of the required bandwidth. Last section deals with a downsampler for decreasing the sampling frequency to its original value. The upsamplig process is made through insertion of one zero between every two input samples. Hence for each output sample only half of the FIR filter coefficients are used. This means that in one time instant the input samples are processed through the even coefficients and next time instant through the odd coefficients of the filters $H(z)$ and $H_{\text{FD}}(z)$. According to this technique and using multirate processing noble identities [23], such processing can be represented as shown in Figure 3, where filters $H_0(z)$ and $H_1(z)$ are the first and second polyphase components of the half band filter $H(z)$. In the same way $H_{\text{FD},0}(z)$ and $H_{\text{FD},1}(z)$ are the first and second polyphase components of fractional delay filter $H_{\text{FD}}(z)$. As can be seen the input sampling frequency is the same for all filters in the resulting structure.

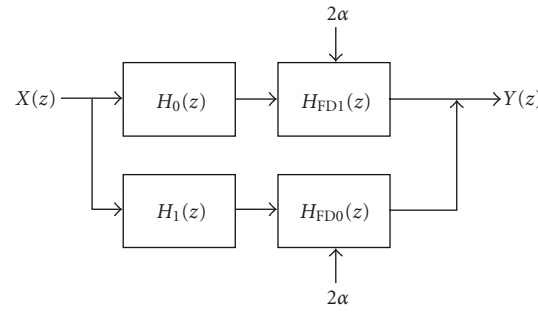


Figure 3: Resulting structure for FD filter.

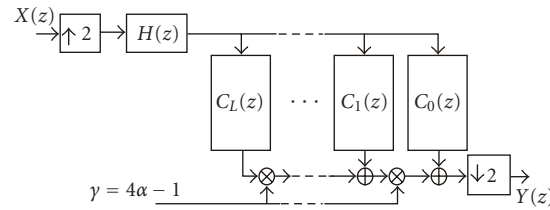


Figure 4: Initial structure of the proposed method.

4. Proposed Design Method

The proposed method for FD filter design with a wide bandwidth and high fractional delay resolution is based on a frequency domain optimization approach, described in the second section, applied to the FD multirate structure, described in last the section.

As mentioned before the maximum frequency of the FD filter in the multirate structure is half of the desired bandwidth. In this way the frequency optimization is made only on the half of the required passband**.** That means that the upper frequency limit in (4) is $\omega_p/2$. This optimization frequency range decrease allows an abrupt coefficient computing time reduction for the wide bandwidth FD filters and the resulting structure requires $C_l(z)$ filters with smaller length N.

In Figure 4 is shown the initial structure, where a double fractional delay value is considered in the update parameter γ, as a result of the doubled processing sampling frequency. The resulting structure after applying noble identities is shown in Figure 5, where the filters $C_{l,0}(z)$ and $C_{l,1}(z)$ are the first and second polyphase components of $C_l(z)$, respectively.

The $H(z)$ filter plays a key role in resulting bandwidth and fractional delay resolution of FD filter. The higher the stopband attenuation of the filter $H(z)$, the higher the resulting fractional delay resolution. Similarly the smaller the transition band of $H(z)$, the higher the resulting bandwidth of the FD filter. Both conditions imply the use of a high-order $H(z)$ filter.

In order to reduce the total number of arithmetic operations per output sample the filter $H(z)$ is designed as

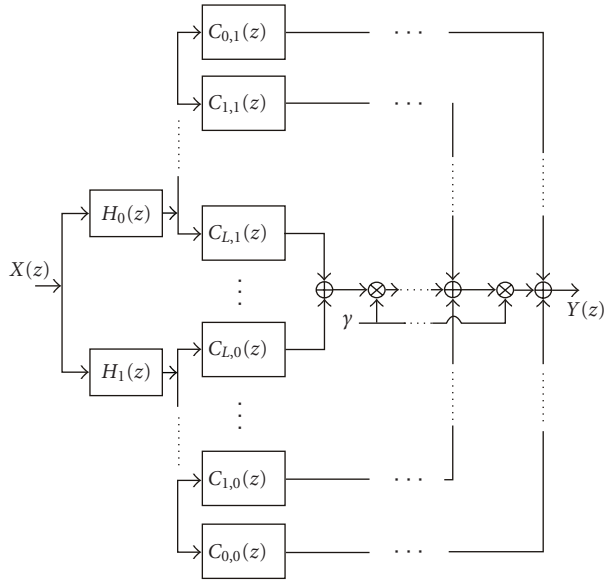


Figure 5: Resulting proposed structure.

a half hand FIR filter. In this way the total number of MPS and APS is given as

$$\begin{aligned} \mathrm{MPS}_2 &= (N)(L+1) + \left\lfloor \frac{N_{LP}+1}{4} \right\rfloor + L, \\ \mathrm{APS}_2 &= (N-1)(L+1) + L + \frac{N_{LP}-1}{2}, \end{aligned} \tag{6}$$

where N_{LP} is the number of coefficients of $H(z)$.

5. Obtained Results

The design method was implemented in MATLAB. An illustrative design example is presented with an FD filter bandwidth of 0.9π and a fractional delay resolution of 1/10000.

The FD filter design using WLS method [20] results in an implementation processing arithmetic of MPS = 703 and APS = 775 with design parameters $N = 87$ and $L = 7$ and weighting functions given by

$$W_1(\omega) = \begin{cases} 1, & \omega \in [0, 0.88\pi), \\ 10, & \omega \in [0.88\pi, 0.8994\pi), \\ 0, & \omega \in [0.8994\pi, \pi], \end{cases} \tag{7}$$

$$W_2(p) = 1.$$

In the same way the resulting processing arithmetic for the implementation of the FD filter example based on variable

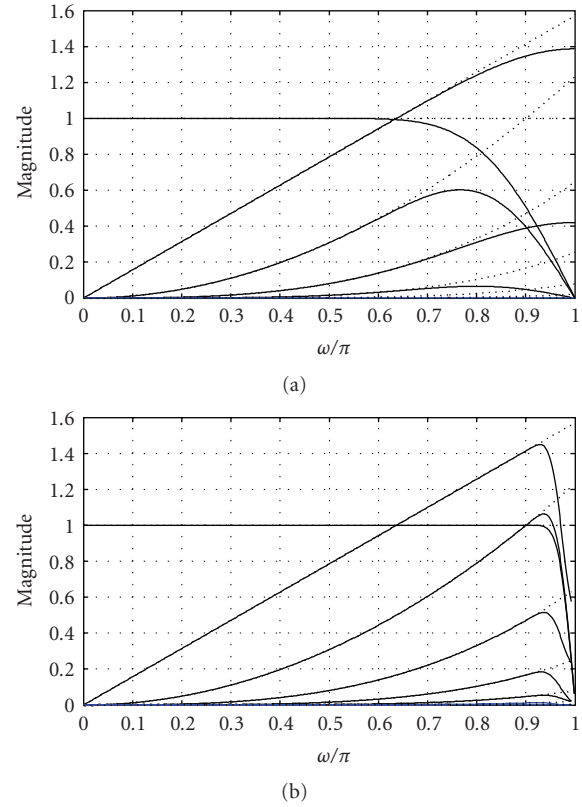


Figure 6: Differentiators ideal responses (dot line) and approximations (solid lines). (a) Proposed method $\omega_p = 0.45\pi$, $N = 12$, and $L = 14$. (b) Direct all band optimization $\omega_p = 0.9\pi$, $N = 104$, and $L = 12$.

FD design method of [21] is MPS = 543 and APS = 535 with design parameters $N = 67$ and $L = 7$ and next weighting function:

$$W_1(\omega) = \begin{cases} 1 & \omega \in [0, 0.88\pi), \\ 3 & \omega \in [0.88\pi, 0.8994\pi), \\ 0 & \omega \in [0.8994\pi, \pi]. \end{cases} \tag{8}$$

The direct use of the frequency domain method [18] with design parameters of $L = 12$ and $N = 104$ results in a total number of MPS = 688 and APS = 1352.

For the proposed design method an interpolator filter $H(z)$ with 241 coefficients was used, designed with a Dolph-Chevishev window having a stopband attenuation of 140 dBs. The design parameters are $L = 12$ and $N = 14$ with a resulting processing arithmetic of MPS = 254 and APS = 242.

In the proposed method the frequency optimization is applied up to $\omega_p = 0.45\pi$ and in the direct method to $\omega_p = 0.9\pi$, as is depicted in Figure 6, where the first seven differentiator approximations are shown for both methods. This half frequency range optimization implies a notable

Table 1: Comparison of approximation errors for several methods.

Methods	$e_{\max}$ (dBs)	e_{rms}
WLS design [19]	−100.0088	2.9107×10^{-6}
Improved WLS [24]	−100.7215	2.7706×10^{-6}
Discretization-free [20]	−99.9208	4.931×10^{-4}
Variable Factional Delay [21]	−99.3669	2.8119×10^{-6}
	−100.9278	2.8497×10^{-6}
Direct Taylor approximation [18]	−93.69	4.81×10^{-4}
Proposed method	−86.17	2.78×10^{-4}

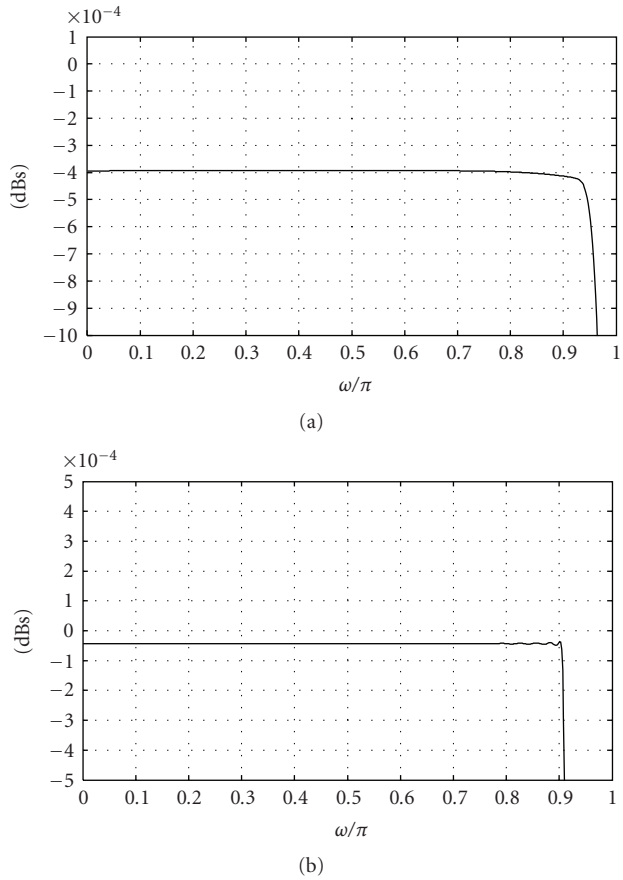


Figure 7: All band magnitude responses for (a) proposed method and (b) direct all band optimization.

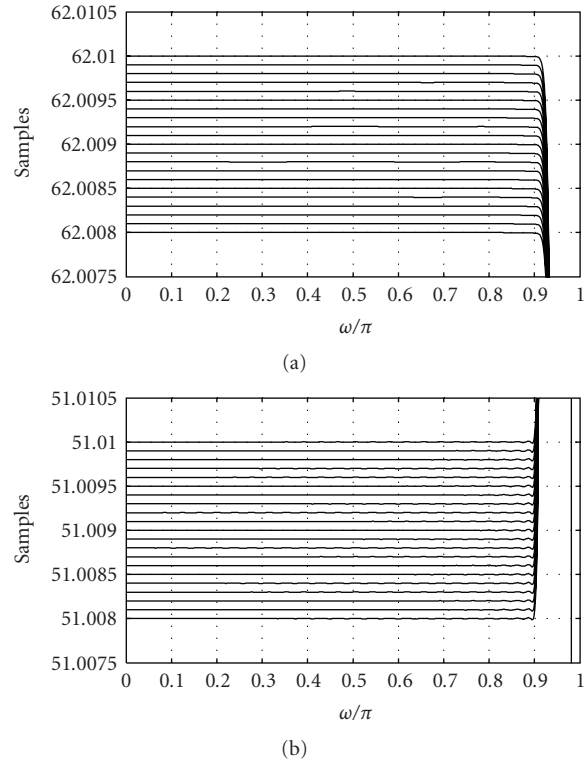


Figure 8: All band fractional delay responses for (a) proposed method and (b) direct all band optimization.

computing workload reduction. The MATLAB computing time in a PC running at 2 GHz for the proposed method is 6.95 seconds and for direct frequency optimization of [18] is 322 seconds.

The all band magnitude responses and group delays for fractional delay values range from 0.0080 to 0.0090 using the direct frequency FD design method and the proposed method results are shown in Figures 7 and 8, respectively.

According to obtained results the proposed method has smaller number of operations per output sample. In order to compare the achieved proposed method approximation with the one obtained with existing methods, the frequency domain error $e(\omega, \alpha)$, the maximum absolute error $e_{\max}$, and the root mean square error e_{rms} are defined, like in [21], as

$$
\begin{aligned}
e(\omega,\alpha) &= |H(\omega,\alpha) - H_d(\omega,\alpha)|, \\
e_{\max} &= \max\left\{e(\omega,\alpha),\ 0 \le \omega \le \omega_p,\ 0 \le \alpha \le 1\right\}, \\
e_{\text{rms}} &= \left[\int_0^{\omega_p}\int_0^1 e^2(\omega,\alpha)\,d\alpha\,d\omega\right]^{1/2},
\end{aligned}
\tag{9}
$$

where $H(\omega, \alpha)$ and $H_d(\omega, \alpha)$ are the frequency response of the designed and ideal FD filters, respectively, and ω_p is passband frequency of the FD filter.

The obtained maximum absolute error and the root mean square error are presented in Table 1; for comparison purpose the results obtained by using the approaches in [18–21, 24] are also presented.

6. Conclusions

A frequency optimization design approach for wide bandwidth and high fractional value resolution FD filters has been proposed. These specifications coupled with the capability of updating the fractional delay value in real-time make the resulting structure suitable to perform important physical layer functions for software-defined radio applications, such as digital symbol synchronization and sample rate conversion.

The obtained results show that the design method notably reduces the coefficients computing workload. The resulting structure allows fractional delay values of 1/10000 of sample and a bandwidth of $0 \leq \omega \leq 0.9\pi$, with a reduced number of arithmetic operations per output sample. The described method is based on a least mean square frequency optimization for coefficients computation. In a future work we will consider the use of other optimization methods in the same proposed approach.

References

[1] J. Mitola, "The software radio architecture," *IEEE Communications Magazine*, vol. 33, no. 5, pp. 26–38, 1995.

[2] J. Vankka, *Digital Synthesizers and Transmitters for Software Radio*, Springer, Dordrecht, The Netherlands, 2005.

[3] T. Hentschel and G. Fettweis, "Sample rate conversion for software radio," *IEEE Communications Magazine*, vol. 38, no. 8, pp. 142–150, 2000.

[4] K. S. Yeung and S. C. Chan, "The design and multiplier-less realization of software radio receivers with reduced system delay," *IEEE Transactions on Circuits and Systems I*, vol. 51, no. 12, pp. 2444–2459, 2004.

[5] V. Lehtinen, D. Babic, and M. Renfors, "On impulse response symmetry of farrow interpolators in rational sample rate conversion," in *Proceedings of the International Symposium on Control, Communications and Signal Processing, (ISCCSP '04)*, pp. 693–696, 2004.

[6] M. Makundi, T. Laakso, and A. Hjorungnes, "Iterative symbol synchronization using variable IIR and FIR fractional-delay filters with arbitrary oversampling ratios," in *Proceedings of the 1st International Symposium on Wireless Communication Systems*, pp. 105–109, September 2004.

[7] S. Pietro, G. Paolo, and F. Lorenzo, "An iterative algorithm for joint symbol timming recovery and equalization of short bursts," *Journal of Communications*, vol. 3, no. 4, pp. 34–40, 2008.

[8] D. Tian-Bo, "Design and parallel implementation of FIR digital filters with simultaneous variable magnitude and non-integer phase-delay," *IEEE Transactions on Circuits and Systems II*, vol. 50, no. 5, pp. 243–250, 2003.

[9] T. I. Laakson, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay: tools for fractional filter design," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60, 1996.

[10] C. W. Farrow, "A continuously variable digital delay element," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 2641–2645, June 1988.

[11] J. Vesma and T. Saramaki, "Optimization and efficient implementation of FIR filters with adjustable fractional delay," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2256–2259, June 1997.

[12] J. Vesma and T. Saramaki, "Design and properties of polynomial-based fractional delay filters," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 104–107, May 2000.

[13] G.-S. Liu, C.-H. Wei, et al., "A new variable fractional sample delay filter with nonlinear interpolation," *IEEE Transactions on Circuits and Systems II*, vol. 39, no. 2, pp. 123–126, 1992.

[14] J. Vesma, *Timing adjustment in digital receivers using interpolation*, M.S. thesis, Department of Information Technology, Tampere University of Technology, Tampere, Finland, 1995.

[15] S. Samadi, M. O. Ahmad, and M. Swamy, "Characterization of B-spline digital filters," *IEEE Transactions on Circuits and Systems I*, vol. 51, no. 4, pp. 808–816, 2004.

[16] C. Cagatay, "An efficient filter structure for lagrange interpolation," *IEEE Signal Processing Letters*, vol. 14, no. 1, pp. 17–19, 2007.

[17] A. Haghparast and V. Välimäki, "A computationally efficient coefficient update technique for lagrange fractional delay filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3737–3740, March 2008.

[18] J. Vesma, R. Hamila, T. Saramaki, and M. Renfors, "Design of polynomial-based interpolation filters based on Taylor series," in *Proceedings of the 10th European Signal Processing Conference*, pp. 283–286, September 1998.

[19] A. Tarczynski, G. D. Cain, E. Hermanovicz, and M. Rojewski, "WLS design of variable frequency response FIR response," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2244–2247, June 1997.

[20] T.-B. Deng, "Discretization-free design of variable fractional-delay FIR digital filters," *IEEE Transactions on Circuits and Systems II*, vol. 48, no. 6, pp. 637–644, 2001.

[21] H. Zhao and J. Yu, "A simple and efficient design of variable fractional delay FIR filters," *IEEE Transactions on Circuits and Systems II*, vol. 53, no. 2, pp. 157–160, 2006.

[22] N. Murphy, A. Krukowski, and I. Kale, "Implementation of a wide-band integer and fractional delay element," *Electronics Letters*, vol. 30, no. 20, pp. 1658–1659, 1994.

[23] P. Vaidyanathan, *Multirate Systems and Filter Banks*, Signal Processing Series, Prentice Hall, Upper Saddle River, NJ, USA, 1993.

[24] W.-S. Lu and T.-B. Deng, "An improved weighted least-squares design for variable fractional delay fir filters," *IEEE Transactions on Circuits and Systems II*, vol. 46, no. 8, pp. 1035–1040, 1999.