Approach for cluster-based spectrum sensing over band-limited reporting channels

Article in IET Communications - October 2012
DOI: 10.1049/iet-com.2010.0510

2 authors:

Xavier N Fernando
Ryerson University
256 PUBLICATIONS 2,146 CITATIONS
SEE PROFILE

Sattar Hussain
Ryerson University
18 PUBLICATIONS 216 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project A Scalable Wireless Positioning Sensor Network for Industrial Automation View project

Electric Vehicles View project
Approach for cluster-based spectrum sensing over band-limited reporting channels

S. Hussain X. Fernando

Department of Electrical and Computer Engineering, Ryerson University, 350 Victoria St., Toronto M2S 2K3, Canada
E-mail: sattar.hussain@ryerson.ca; fernando@ee.ryerson.ca

Abstract: In this study, the authors address the problem of bandwidth limitations of the reporting channels in cognitive radio (CR) networks. They propose a cluster-based spectrum-sensing approach that minimizes the bandwidth requirements by reducing the number of terminals reporting to the fusion centre to a minimal reporting set. The approach replaces the secondary base station by a local fusion centre and combats the destructive channel conditions by replacing the global reporting channels with local ones. They also propose a new approach to select the location of the local fusion centre using the general centre scheme in graph theory. The minimal dominating set (MDS) clustering algorithm is used to obtain the minimal set of clusters that keep the network connected. This study investigates how the sensing efficiency, the sensing accuracy, and the per-node throughput are affected by the cluster size, the number of clusters, and the reporting channels error. The results obtained reveal that the cluster-based cooperative sensing system outperforms the conventional cooperative sensing system in terms of throughout capacity especially when the reporting channels are subjected to a high probability of error. A systematic way to find the optimal number of cooperative clusters that gives a minimum probability of false alarm is presented.

1 Introduction

Ideally, secondary users should guarantee no interference to the primary user. However, in a real communication environment, the communication channels between the primary and secondary users may be impaired because of shadowing and fading. Under deep shadowing/fading, cognitive radio (CR) users may mistake a weak primary signal for a vacant channel. Therefore individual local sensing is prone to errors and unreliable. This can be improved by sharing the local observations among a few CR users in a cooperative manner adding diversity.

Cooperative spectrum sensing, therefore aims to improve the reliability of primary user detection by exploiting the spatial diversity among cooperative secondary radios. Cooperative sensing, however, requires secondary users to exchange their local observations with a fusion centre (secondary base station) for decision making. For this purpose, reporting channels are used by the CR terminals to report their local observations or individual decisions to the fusion centre. Typically, these channels are band limited and may experience shadowing and multi-path fading. The number of cooperative users and the amount of information that each user should send to the fusion centre determine the required reporting bandwidth over these channels. For a large network, the required bandwidth may exceed the capability of the reporting channels, especially when users send their local statistics (the real-time measurements) instead of their individual decisions.

This issue can be improved by forming clusters to share the sensing information locally [1]. In this approach, the secondary users are grouped into separate clusters. Each group then elects a ‘cluster head’ that locally coordinates the access to the shared spectrum by cluster members. In addition, the cluster head acts as a local fusion centre to forward an aggregated decision (hard decision fusion) or a linear combination of the local observations (soft decision fusion) to the fusion centre. With such a scenario, the access to the reporting channel is significantly reduced whereas all the CR users are still taking part in the cooperative process.

Clustering is used widely as a hierarchical approach of topology management in ad hoc wireless networks [2–5]. However, in CR networks, clustering must consider the fact that the available channel sets are changing temporally and spatially. Furthermore, the cluster sizes and cluster locations may change when a CRs neighbourhood changes with each new operating frequency. Consequently, clustering in CR networks is performed according to ‘channels topology’ instead of ‘nodes topology’. Specifically, a CR node forms a cluster on an available channel and invites the adjacent nodes to join its cluster if the same channel is available in their channel sets [6].

Cluster-based spectrum sensing is divided into rounds of three phases: sensing, cluster setup and transmission phase. During the sensing period, spectrum holes are located and channel sets are made available for data exchange. CR users need to synchronise their spectrum sensing phase to avoid false alarms that may be triggered by some CRs that started their spectrum sensing earlier [7]. In the second phase, clusters are formed and cluster heads are elected. Finally, the transmission phase starts when the CR
terminals communicate and exchange data using the set of the sensed channels that are originally owned by the primary network. The available band is only utilised during this phase and it stays idle during the sensing and clustering phases. The length of the transmission period determines how efficiently the available spectrum band can be utilised.

Thus, even though the cluster setup time is much shorter than the channel transmission time, it is not preferable to perform the cluster setup at every sensing round so as to improve sensing efficiency. Fig. 1 illustrates the sensing round structure for a cluster-based cooperative sensing.

Network traffic in cluster-based systems is generated mainly by intra-cluster and inter-cluster communication. Inter-cluster communication happens between the cluster heads and the traffic relay gateways. A gateway is a CR user who might be in one hop from two neighbouring cluster heads in case of overlapping clusters or in one hop to another gateway in an adjacent cluster in the case of disjoint clusters. In both traffic types, the packets generated by a source node may reach the destination node through a single-hop or a multi-hop routing. For clustering to be effective, the number of cluster heads, gateways and the links that are connecting them must be minimised while preserving the connectivity of the whole network [8]. Minimising the number of clusters reduces the overhead traffic and the network maintenance requirements. However, clustering must consider the traffic load at both cluster heads and the gateways as these nodes tend to be the bottlenecks of the entire network. Setting up an upper bound for the cluster size is essential here as it prevents the overcrowding of nodes in the clusters and avoids traffic congestion at the interconnection nodes.

In this paper, we investigate the performance of expanded minimal dominating set (MDS)-based clustering algorithm scheme and introduce an approach to select the best location for the local fusion centre that acts as a master fusion centre in a totally distributed system. The clustering algorithm aims to find the minimal number of clusters that cover all the users operating in the field while preserving network connectivity. The suggested approach aims to reduce the bandwidth required for reporting data to the fusion centre by reducing the number of reporting terminals. We also study how the sensing efficiency, sensing accuracy and the per-node throughput are effected by the various parameters of the clustering scheme.

2 Network model

In this work, we assume that one of the elected cluster heads acts as a centre entity and serves as a master fusion centre to finalise the cooperative decision and broadcast it back to the other cluster heads and from there to the cluster members. The proposed network scheme has two levels of data fusion (data aggregation). The first data fusion is performed at the cluster level by each cluster head and the second one is performed at the master fusion centre. We assume that NC CR terminals are distributed in a square field of area A as a two-dimensional Poisson-point process with a density λ. A node (CR terminal) v is said to be in the neighbourhood of a node u, if v are within a distance at most rc from u, where rc is the cluster radius. Each CR terminal is identified by a unique identification (ID) and the CR terminals are assumed static or moving slowly during the algorithm execution.

The topology of the CR network is presented by an undirected graph G(V, E), where V is a set of vertices in this graph that stands for CR terminals and E is a set of links between those terminals. A link (u, v) ∈ E means that terminals u and v share at least one common channel and in one-hop distance from each other. The neighbourhood set of a given node v ∈ V, represents all the nodes in a one hop away from v that share at least one channel with v.

Transmission power determines the cluster radius and largely affects the number of clusters, cluster size and network performance. Higher transmission power means fewer hops resulting in higher network throughput. However, the higher interference resulting from higher transmission power tends to limit the network throughput. This is true in the other way as well.

This work investigates the impact of transmission power on the network topology. Basically, we investigate how the average number of isolated nodes (Niso) is changing with cluster radius. Isolated nodes refer to those nodes that cannot find any neighbouring terminal within a radius of rc in any channel within their channel sets. These nodes declare themselves as cluster heads and form their own clusters (single-node clusters).

As CR terminals are distributed as a point Poisson process with a density λ, then the probability that a node is unable to find any neighbouring node within a radius rc = e\(^{-λπr_c^2}\). If we consider a two-dimensional polar coordinate (r,θ), where the node is in the origin, then the average of isolated node E(Niso)

![Fig. 1 Sensing round structure in cluster-based spectrum sensing](image-url)
can be calculated as follows (see [9])

\[ E(N_{iso}) = \lambda A \int_{0}^{\infty} e^{-\lambda \pi r^2} \lambda r \, dr \, d\theta \]  
\[ E(N_{iso}) = \lambda a e^{-\lambda \pi r^2} \]  

(1)

where, \( h_p \) denotes the number of transmission hops. As our approach is to establish a connected MDS of connected clusters that covers all the deployed CR terminals, then the fraction of single-node clusters, \( E(N_{iso}) \), must be very small. If \( \kappa, (\kappa > 0) \), is an arbitrarily small number that represents a target percentage of single-node clusters such that

\[ E(N_{iso}) \leq \kappa \lambda A \]  
(2)

then by equating (2) and (3) we obtain the following lower bound of cluster radius at this specified target

\[ r_c \geq \frac{-\log \kappa}{\pi h_p^2} \]  
(3)

Increasing the cluster radius will clearly increase the probability to find a neighbouring terminal for each node and increase the cluster size. However, as the cluster size increases, the probability to have network bottlenecks at the inter-cluster communication gateways increases. Also, the inter-cluster and intra-cluster interferences tend to increase with larger cluster communication range. Hence, to avoid congestion at those nodes, we introduce an upper bound for the cluster size such that the traffic load at those gateways will be upper bounded and adhere with the quality of service (QoS) requirements.

With low cross correlation code division multiple access (CDMA) codes, inter-cluster interference can be eliminated. We assume each cluster is assigned a unique transmitting code that is different from those codes used in neighbouring clusters. As the receiver nodes must be set to the same code as the designated transmitter, interference with other clusters is avoided. If no two nodes in a cluster are transmitting simultaneously, there will be no intra-cluster interference. Following [10], we assume that within each cluster, the channel is slot synchronised using time division multiple access (TDMA) scheme in which each node is assigned a single time slot for transmission.

### 3 Clustering algorithm

In graph theory [8], the MDS problem is to find a subset \( \mathcal{C}(|\mathcal{C}| = M) \) called the dominating set such that each node belongs to one member of \( \mathcal{C} \). We refer to the dominating set members as cluster heads and the nodes that belong to one of the cluster heads as cluster members. Finding an MDS is NP-hard in general [5, 11], but sufficient approximation algorithms exist. These algorithms obtain a sub-optimum dominating set by generating a local minimum election of the dominators.

The algorithm requires a preliminary node discovery and a master channel configuration phase. The master channel is used by the cluster heads and cluster members to exchange the control data necessary for cluster formation. We assume that at the completion of the node discovery, any node say \( v_i \) knows the channel availability set of node \( v_j \in v_i^l \) and the set of one-hop neighbours of the node \( v_j \), \( \forall v_j \in v_i^l \). Where \( v_i^l \) denotes the one-hop neighbourhood of node \( v_i \).

The MDS cluster setup scheme is performed after the preliminary setup phase. As the network topology needs to be optimised from time to time as the network conditions changes, the MDS scheme is required to be performed periodically. Typically network topology changes when a primary user appears at a certain channel, a new terminal joins/leaves the network, or a cluster head reporting channel comes under a deep shadowing/fading.

In MDS scheme, every node \( i \in V \) is required to be covered by one member of the dominating set \( \mathcal{C} \subseteq V \). The dominating set contains \( M \) subsets \( \mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_M \) of the base set \( V = \{1, 2, \ldots, N\} \) such that \( \cup_{j=1}^{M} \mathcal{C}_j = V \). We define a binary variable \( x_j \) for the subsets \( \mathcal{C}_j \), \( j = 1, 2, \ldots, M \) as follows

\[ x_j = \begin{cases} 
1 & \text{if } \mathcal{C}_j \subseteq \mathcal{C} \\
0 & \text{otherwise}
\end{cases} \]

By defining \( a_{ij} \) to be 1 when a node \( i \in \mathcal{C}_j \) and 0 otherwise, we can write the problem as

\[ \min \sum_{j \in \mathcal{C}} x_j \quad j = 1, 2, \ldots, M \]

subjected to

\[ \forall i \in V, \sum_{j \in \mathcal{C}} a_{ij} = 1 \]

\[ \forall i \in V, \sum_{j \in \mathcal{C}} a_{ij} a_{il} \leq h_p r_c \]

\[ \forall j \in \mathcal{C}, \sum_{i \in \mathcal{C}} a_{ij} \leq P \]

where, \( P \) is an upper bound of cluster size.

We extend the single-hop CogMesh algorithm presented in [6] to a \( d \)-hop clustering scheme and generalise this algorithm for the above described system model. The MDS algorithm is performed at the node level to minimise the number of clusters in the neighbourhood of a selected node and reconfigure the whole cluster topology when a new cluster set smaller than the original one can be found. Instead of selecting a node in a random way as in [6], our approach selects a node with the highest reporting channel gain from a set of neighbouring nodes that share a channel with a maximum degree to be a cluster head. The algorithm starts forming a cluster from this cluster head and all nodes within \( d \)-hop, \( d = 1, 2, \ldots, h_p \), that share the same channel. The new cluster head and all the assigned cluster members will be eliminated from the node neighbourhood set. Then, a node with a maximum degree on another channel will be selected as a cluster head and its assigned cluster members eliminated too from the remaining node set and so on until all the nodes are configured in the new cluster topology. The algorithm then starts the gateway nodes selection to construct the inter-cluster communication. A priority case is used in gateways selection. A node in the shortest path between any two cluster heads is given the highest priority. Once a cluster is formed, the cluster head communicates with the neighbours to select the CDMA codes. Only when the code assignment is completed, data can be transmitted in the network.
Lemma 1: In a $d$-hop cluster, the distance between any two

Prove: Let $v$ be any node belongs to cluster $C_j, j = 1, 2, \ldots, M$ whose cluster head is $m_j$. If $v(\text{ID}) \neq m_j(\text{ID})$ then there must be another node $u \in C_j$ at most in a $d$-hop away from $v$ where $u(\text{ID}) = m_j(\text{ID})$. Hence, $u$ is the cluster head of the cluster $C_j$ and $|v - u| \leq drc$. For any other node say $w \in C_j$, whose $w(\text{ID}) \neq m_j(\text{ID})$, $|w - u| \leq drc$. Therefore the distance between $v$ and $w$ is at most $2drc$.

Lemma 2: If $C$ is a connected dominating set, then any cluster head is at most $(2d + 1)r_c$ away from the nearest cluster head.

Prove: Let us assume the contrary and assume that the closest routing path between cluster heads $m_1 \in C_1$ and $m_2 \in C_2$ is $(2d + 2)r_c$. Let this closest path pass through the two gateways, $x$ and $y$ (see Fig. 2). As $x \in C_1$ and $y \in C_2$, then $|m_1 - x| \leq dr_c$ and $|m_2 - y| \leq dr_c$, and $|x - y| \geq (2d + 2)r_c - 2dr_c = 2r_c$. This implies that each of $x$ and $y$ is not in a one-hop transmission range of each other and the two clusters are not connected. As the dominating set is a connected dominating set, then, there must be another gateway $v \in C_1$ which is in one hop to another cluster head say, $m_1 \in C_1$ or adjacent node say $z \in C_1$ which is at most $d$-hop from $m_3$, that is, $|m_3 - z| \leq dr_c$. Therefore $|m_1 - m_3| = |m_1 - v| + |v - z| + |z - m_3| \leq dr_c + r_c + dr_c = (2d + 1)r_c$ and $m_3$ is the nearest cluster head to $m_1$.

The master fusion centre is then elected from the minimal cluster head set. The placement of the fusion centre also needs to be optimised. The fusion centre placement problem is concerned with selecting the best location in a specified region for the network centre entity. Mainly, there are two options to solve this location problem, the centre problem and the general centre [12]. In graph theory, the graph centre is any vertex $v$ whose furthest vertex is as close as possible whereas the general centre is any vertex $v$ where the aggregated distance from all other vertices is as minimum as possible. Let

$$Mv(i) = \max_{j \neq i} d_{ij}$$

denotes the maximum distance of any vertex from vertex $i$, where $d_{ij}$ is the vertex-to-vertex distance and

$$Sv(i) = \sum_{j \neq i} d_{ij}$$

denotes the aggregated distance of all vertices from vertex $i$. Then the graph centre is any vertex $x$ such that

$$Mv(x) = \min_{i} \{Mv(i)\}$$

whereas the general centre is any vertex $x$ with the smallest possible $Sv(i)$, that is, the smallest aggregated distance from all other vertices or

$$Sv(x) = \min_{i} \{Sv(i)\}$$

Practically, the reporting paths are subjected to different channel conditions. Therefore assuming identical channel conditions for all the reporting paths is not realistic. Moreover, in CR networks, priority is given to detection accuracy which is strongly dependent on the channel gain of each path. Then for a cooperative system, the shortest reporting channels may not always be the best choice especially if they come under a deep fading/shadowing. In this work, we proposed a modified general centre scheme that considers the channel gain of the $M - 1$ signal paths between the fusion centre and the $M - 1$ cluster heads. In the proposed scheme, the fusion centre is the cluster head whose aggregated channel gain has the maximum possible value. As the cooperative decision is made by combined received signals from different cluster heads, we believe that the channel gain is the best parameter to be considered for this problem. Moreover, the channel gain is affected directly by both the transmitter–receiver distance and the random influence of fading and shadowing conditions.

In a fading and shadowing environment, the channel gain between a transmitter, $i$, and a receiver, $j$, is modelled by [13]

$$G_{ij} = K \frac{10^{n/10}}{d_{ij}^a}$$

where, $K$ is a constant depending on carrier frequency, antenna height and antenna gain, $d_{ij}$ denotes the distance between the transmitter and the receiver, $\alpha$ denotes the path loss exponent $2 < \alpha < 6$ and $\beta$ is a zero mean Gaussian random variable with variance $\sigma^2$. In practice, $5 < \beta < 12$ and $10^{n/10}$ represents the shadowing factor with a log-normal distribution [14, 15]. Log-normal shadowing is usually characterised in terms of its dB-spread and indicates how the loss in dB varies about its mean value.

Clearly, the gain is affected directly by the two factors, the distances between nodes and the random influence of fading and shadowing. Thus, using the gain model in finding the best location of the fusion centre is more accurate than considering merely the topological distance. Accordingly, the master fusion centre is the cluster head $m$ that has the maximum aggregate channel gain or

$$G_{hh}(m) = \max_{i} \{G_{hh}(i)\} \quad i = 1, 2, \ldots, M$$

Fig. 2 Maximum distance to the nearest cluster head
where

\[
Gbh(i) = \sum_j G_{ij} \quad j = 1, 2, \ldots, M
\]

4 Spectrum-sensing model

The energy detector [16], which is a simple non-coherent suboptimal detector, is widely considered for local spectrum sensing in CR networks. For the i-th user, the objective of local sensing, is to decide between two hypotheses, \(H_0\), for primary user absence and \(H_1\), for primary user presence.

\[
\begin{align*}
H_0: & \quad y_i(t) = n_i(t) \\
H_1: & \quad y_i(t) = h_i(t)s(t) + n_i(t)
\end{align*}
\]

(8)

where \(y_i(t)\) is the signal received at the i-th receiver, \(s(t)\) is the signal transmitted by the primary user, \(h_i(t)\) is the channel gain assumed to be constant during the detection interval, and \(n_i(t) \sim N(0, \sigma_n^2)\) is an additive white Gaussian noise with zero mean and variance, \(\sigma_n^2\). \(s(t)\) and \(n(t)\) are assumed to be statistically independent.

Over a sensing time window, \(T_s\), the CRs collect a test statistics, \(Y\), and the decision rule at each radio is given by

\[
Y \overset{H_1}{\cong} Y \overset{H_0}{\cong} \varepsilon
\]

(9)

where, \(\varepsilon\) is the corresponding decision threshold. Therefore the probability of false alarm, \(P_f\), and the probability of detection, \(P_d\), at the i-th secondary radio can be defined as

\[
P_f = P(Y_i > \varepsilon|H_0)
\]

(10)

\[
P_d = P(Y_i > \varepsilon|H_1)
\]

(11)

In a non-fading environment where \(h\) is deterministic, the exact closed form of \(P_f\) and \(P_d\) expressions is given by [16]

\[
P_f = \frac{\Gamma(N, \varepsilon/2)}{\Gamma(u)}
\]

(12)

\[
P_d = Q_u\left(\sqrt{2\gamma_i}, \sqrt{\varepsilon}\right)
\]

(13)

and the probability of missed detection, \(P_m\)

\[
P_m = 1 - P_d
\]

(14)

where, \(\gamma_i\) denotes the received signal-to-noise ratio (SNR) at the i-th radio, \(u = T_s/\text{BW}\) is the time bandwidth product assumed to be an integer number representing number of the collected samples, and \(\Gamma(.)\) and \(\Gamma(\cdot, \cdot)\) are the complete and incomplete \(\Gamma\) functions, respectively.

Under Rayleigh fading, the SNR, \(\gamma_i\), follows an exponential distribution. Therefore assigning a constant SNR to each transmission path becomes unrealistic. Average SNR, \(\overline{\gamma}\), is a more appropriate performance measure than instantaneous SNR in fading channel. Accordingly, the closed-form expression of \(P_d\) is given by [17]

\[
P_d = e^{-\left(\varepsilon/\sqrt{2}\right)} \sum_{k=0}^{u-1} \frac{1}{k!} \left(\frac{\varepsilon}{\sqrt{2}}\right)^k + \frac{1 + \gamma_{\overline{\gamma}}}{\gamma_{\overline{\gamma}}} e^{-\frac{1}{\gamma_{\overline{\gamma}}}} e^{-\gamma_{\overline{\gamma}}/2} \sum_{k=0}^{u-1} \frac{1}{k!} \left(\frac{\gamma_{\overline{\gamma}}}{2(1 + \gamma_{\overline{\gamma}})}\right)^k
\]

(15)

5 Cluster-based spectrum sensing

To improve the spectrum sensing, the MDS-based clustering scheme introduced in Section 3 is used to allow the secondary users in each cluster to cooperatively share their local observations. The sensing information is shared only at the local scale within each cluster and there is no need for each CR terminal to send its own decision to the master fusion centre.

For a cluster \(C_j\), with a cluster size of \(N_{C_j}\), the individual local sensing statistics of cluster members \(Y_i, i = 1, 2, \ldots, N_{C_j}\), are linearly combined at the cluster head

\[
Y_j = \sum_{i=1}^{N_{C_j}} Y_i \quad \forall C_j, j = 1, 2, \ldots, M
\]

(16)

As \(Y_j\)s are normal random variables, according to (16), their linear combination is also normally distributed with the following mean values and variances

\[
Y_j \sim N\left(u \sum_{i=1}^{N_{C_j}} \sigma_i^2, 2u \sum_{i=1}^{N_{C_j}} \sigma_i^4\right)
\]

\[
H_0
\]

\[
H_1
\]

The cluster head then makes the cluster decision by comparing the linear combination with a pre-fixed threshold, \(\varepsilon_{Cj}\). An approximate-form expression for \(P_f\) and \(P_d\) computed at each cluster head can be defined by

\[
P_{fC} = Q \left[\frac{\varepsilon_{Cj} - u \sum_{i=1}^{N_{C_j}} \sigma_i^2}{\sqrt{2u} \sum_{i=1}^{N_{C_j}} \sigma_i^4}\right]
\]

(17)

\[
P_{dC} = Q \left[\frac{\varepsilon_{Cj} - u \sum_{i=1}^{N_{C_j}} (\gamma_i + 1)\sigma_i^2}{\sqrt{2u} \sum_{i=1}^{N_{C_j}} (2\gamma_i + 1)\sigma_i^4}\right]
\]

(18)

Equations (17) and (18) clearly illustrate that the spectrum-sensing performance of each cluster is largely effected by the number of CR terminals in each cluster (cluster-size), the threshold \(\varepsilon_{Cj}\) and the number of samples collected by the detector during the sensing time. In fact, \(P_f\) refers to the probability of white spaces that are misclassified as occupied and \(P_m\) refers to the probability of harmful interference to the primary user. From the secondary user’s point of view, a lower probability of false alarm, means more spectrum access opportunities. However, from the primary user’s side, the higher probability of detection means more protection from harmful interference. Depending on which probability is of interest, that one is fixed and the other one is
optimised. As the priority is given to interference avoidance to the primary network, a high detection probability must be guaranteed. Therefore if we set a lower bound for the detection probability say $P_{dc}$ at the cluster level, then the objective is to minimise the probability of false alarm as much as possible so as to increase the spectrum access opportunities and thereafter the network throughput. By combining (17) and (18)

$$P_{fc} = \frac{Q(\hat{P}_{dc})^{-1}}{4u} \left[ \frac{2\sum_{j=N}^{\infty} + (1 + 2\gamma) \frac{\sigma_i^n}{\sum_{i=1}^{\infty} \alpha_i^n} |h_i|^2}{\sqrt{2u} \sum_{i=1}^{\infty} \sigma_i^n} \right]$$

(19)

Under the bandwidth constraints of the reporting channel, we allow each cluster head to send only 1-bit decision $[0]$ for $H_0$ and $[1]$ for $H_1$ to the master fusion centre rather than their decision statistics. The fusion centre then makes the final decision according to the fusion rule implemented. The OR-rule is implemented in the proposed scheme. For the OR-rule, a decision of $[0]$ for $H_0$ is only made when all the $M$ cluster decisions demonstrate the absence of the primary user. Such kind of rule is perfect for interference avoidance to the primary user as the secondary user will only be allowed to access the spectrum if all the cluster heads reported the binary decision $[0]$ to the fusion centre. The decision rule can be defined as

$$D = \begin{cases} 1 & \sum_{j=1}^{M} B_j \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

(20)

Therefore under the OR-rule, the false-alarm probability and the detection probability can be defined as

$$Q_f = 1 - \prod_{j=1}^{M} (1 - P_{fc}^j)$$

(21)

$$Q_d = 1 - \prod_{j=1}^{M} (1 - P_{dc}^j)$$

(22)

A more general formulation of the data fusion problem is introduced in [18] and [19].

Under deep fading/shadowing, a reported decision of $[0]$ for primary network absence may be received at the fusion centre as $[1]$, or a reported decision of $[1]$ for active primary network may be received at the fusion centre as $[0]$. In the first case, a false alarm will be triggered and a missed detection risk may be encountered in the second case. If $\hat{P}_{fc}^j$ denotes the probability of receiving $[1]$ at the master fusion centre when the $j$th cluster head reports $[0]$ and $P_{mc}^j$ denotes the probability of receiving $[0]$ at the master fusion centre when the $j$th cluster head reports $[1]$, then under OR-rule, the probabilities of false alarm, $Q_f$, and missed detection, $Q_m$, become [20]

$$Q_f = 1 - \prod_{j=1}^{M} \left[ (1 - P_{fc}^j)(1 - \hat{P}_{fc}^j) + \hat{P}_{fc}^j \hat{P}_{mc}^j \right]$$

(23)

$$Q_m = \prod_{j=1}^{M} \left[ P_{mc}^j(1 - \hat{P}_{fc}^j) + (1 - P_{mc}^j)\hat{P}_{mc}^j \right]$$

(24)

Let $P_f^j = \hat{P}_{fc}^j = \hat{P}_{mc}^j$, where $P_f^j$ denotes the probability of the reporting errors for cluster head $m_j$, then (23) and (24), can be rewritten as

$$Q_f = 1 - \prod_{j=1}^{M} \left[ (1 - P_{fc}^j)(1 - P_f^j) + P_{fc}^j P_f^j \right]$$

(25)

$$Q_m = \prod_{j=1}^{M} \left[ P_f^j(1 - P_f^j) + (1 - P_f^j)P_f^j \right]$$

(26)

6 Performance evaluation

In this section, we present a computer simulation to demonstrate the performance of the proposed cluster-based system. We assumed a CR network of a 100 terminals deployed in an area of 100 m $\times$ 100 m. The additive noise is assumed as a zero-mean real-valued Gaussian process. The sensing frame, $T_s$ is assumed to be fixed and assigned a value of 20 ms. The same power is assumed to be adopted by each terminal for intra-cluster communication with the
power control capability that is required for the communication between cluster heads and fusion centre.

We conducted several trials with random deployment of CR terminals. In each trial, the cluster radius varied and the percentage of the disconnected node is recorded. An average number of disconnected nodes over these trials is illustrated in Fig. 3 for one-, two- and three-hop clustering scheme, respectively. In Fig. 4, the percentage of the disconnected nodes is plotted against the cluster radius with different node densities for a single-hop clustering scheme.

In Fig. 5, we plot the complementary receiver operating characteristic (ROC) curves for SNR of 0, 5, 10, 15 dB and $u = 6$ samples. The sensing channel is assumed to be
subjected to Rayleigh fading. Obviously, the spectrum sensing performance deteriorates at lower SNR values as illustrated in Fig. 5. Therefore individual sensing becomes unreliable at weak primary signals which is typical in a fading environment.

To demonstrate the effect of the cluster size \( N_c \), on the overall performance of the clustered network, the complementary ROC curves are plotted in Fig. 6. The curves show the probability of missed detection against the probability of false alarm for a cluster size of 5, 10 and 15 nodes. The reporting channel conditions are realised with probability of error, \( P_e = 0.0001 \). Clearly, the network performance is improved by increasing the number of nodes in each cluster. However, the cluster size must be upper bounded to prevent network bottlenecks resulting from overcrowded clusters.

To investigate the throughput performance of the proposed system, a target detection probability of 0.9 is used with a mean value of 0.7 and 0.3 for the idle and the active periods, respectively. Fig. 7 shows that probability of false alarm, \( Q_f \), initially decreases rapidly when the number of cooperative clusters increases. However, it increases later with more cluster heads involved in the cooperative process. In fact, after the initial drop, \( Q_f \) increases with a rate that greatly depends on the probability of error. As the OR-rule is increasing function in terms of \( m \), the initial drop occurs as a result of the term \( Q^{-1}(1 - \sqrt{1 - P(e)}) \) in (25). However, when the probability of false alarm, \( P_f \), becomes very small compared to the probability of errors, \( P_e \), (25) can be reduced to \( Q_f = 1 - \prod_{j=1}^{M} (1 - P_e^j \) and that explains why \( Q_f \) becomes mainly dependent on \( P_e \). An optimal number of cluster heads that give a minimum probability of false alarm is obtained which varies according to the reporting channel probability of error.

In Fig. 8 the normalised per-node throughput is plotted against the number of clusters for reporting channel probability of error, \( P_e = 0.1, 0.01, 0.001 \) and 0.0001. The perfect reporting channel \( (P_e = 0) \) is also plotted for comparison. The figure reveals that at a higher probability of error, the throughput performance deteriorates with more users added to the cooperative decision. This can be easily understood from Fig. 7 where the probability of false alarm increases more rapidly at a higher probability of error.

To compare the cluster-based system with the conventional system, the per-node throughput performance is plotted in Fig. 9 for \( P_e = 0.1 \). Clearly, the cluster-based system outperforms the non-cluster system especially when more users are engaged in the cooperative process. However, the throughput performance for the cluster-based approach is close to that of the non-cluster approach at the optimal value. This can be explained by the fact that the number of cooperative users where the optimal throughput happened is six users only. When such a small number is deployed in a field, most of the clusters are expected to be single-node clusters and consequently the obtained results come close to that of the non-clustered one. However, we must consider the fact that the number of deployed secondary users will not necessarily be at the optimal value all the times. Therefore when the number of cooperative users increases, a significant improvement is obtained as compared to the non-cluster approach as can be seen from Fig. 9. It is worth noting that by adjusting the cluster radius, the number of clusters in the minimal cluster set can be adjusted to match the requirement of both the reporting channel bandwidth and the optimal achievable per-node throughput. Fig. 10 shows that an MDS of six clusters is achieved with a cluster radius of 20 and 12 m for one- and two-hop clustering scheme, respectively.

7 Conclusion

We investigated a distributed spectrum sensing scheme that eliminates the need for a base station and replace it with a local master fusion centre. The MDS approach is used as a clustering scheme and the general centre problem scheme is used as the base to select the master fusion centre. A lower bound of the cluster radius that keeps the number of isolated nodes under an upper limit is determined in this work. The influence of the cluster size, number of cluster, sensing time and the probability of reporting channel errors on the per-node throughput capacity is investigated. The results obtained reveal that under bad channel conditions, it is not necessary to include all the cooperative users for the best performance. Instead, an optimal number of clusters that gives a minimum probability of false alarm and consequently a maximum per-node throughput is obtained. When this optimal number matches the MDS, the reporting channel bandwidth requirement is also achieved.

8 References