

## Energy-efficient spectrum-aware clustering for cognitive radio sensor networks

ZHANG HuaZi<sup>1,2</sup>, ZHANG ZhaoYang<sup>1,2\*</sup> & YUEN Chau<sup>3</sup>

<sup>1</sup> Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China;

<sup>2</sup> Zhejiang Provincial Key Laboratory of Information Network Technology, Hangzhou 310027, China;

<sup>3</sup> Singapore University of Technology and Design, Singapore

Received February 2, 2012; accepted April 20, 2012; published online July 2, 2012

As a smart combination of cognitive radio networks and wireless sensor networks, recently introduced cognitive radio sensor network (CRSN) poses new challenges to the design of topology maintenance techniques for dynamic primary-user activities. This paper aims to provide a solution to the energy-efficient spectrum-aware CRSN clustering problem. Specifically, we design the clustered structure, establish a network-wide energy consumption model and determine the optimal number of clusters. We then employ the ideas from constrained clustering and propose both a centralized spectrum-aware clustering algorithm and a distributed spectrum-aware clustering (DSAC) protocol. Through extensive simulations, we demonstrate that DSAC can effectively form clusters under a dynamic spectrum-aware constraint. Moreover, DSAC exhibits preferable scalability and stability with its low complexity and quick convergence under dynamic spectrum variation.

**cognitive radio sensor network, constrained clustering, energy efficient, distributed spectrum-aware clustering**

**Citation:** Zhang H Z, Zhang Z Y, Yuen C. Energy-efficient spectrum-aware clustering for cognitive radio sensor networks. *Chin Sci Bull*, 2012, 57: 3731–3739, doi: 10.1007/s11434-012-5254-4

As a candidate for next-generation communication networks [1], cognitive radio sensor network (CRSN) [2] has attracted worldwide attention. The dynamic virtue of such network poses new technical challenges on the design of its routing algorithm. Existing cluster-based routing schemes for cognitive radio networks aim to provide joint spectrum and routing decisions, but do not consider CRSN's problem of energy scarcity inherited from traditional wireless sensor networks. Research work [3] has aimed to minimize the number of clusters in a cognitive mesh network while ensuring the connectivity of all users. Distributed clustering for cognitive radio ad hoc networks, in which a clustered structure is established according to the local control channel, has been proposed in [4].

By contrast, clustering schemes for non-cognitive WSNs are developed with the main objective of minimizing energy

consumption. However, they are not designed to deal with spectrum-aware communications. In pioneering work [5], an energy-efficient low-energy adaptive clustering hierarchy protocol was proposed. The hybrid, energy-efficient, distributed clustering approach has been developed [6] to cluster ad hoc sensor networks. Recently, centralized clustering schemes, such as  $k$ -means clustering [7], have been developed and found to further improve the network performance in terms of energy efficiency.

The novelty of our proposed clustered structure is two-fold. On the one hand, the structure is aware of the radio environment. On the other hand, the structure should be energy efficient.

The basic assumptions are listed as follows:

(i) Spectrum-sensing capability: possessing spectrum-sensing capability, each CRSN node can correctly determine the available channels at its location.

(ii) Spectrum-aware constraint: CRSN nodes that belong

\*Corresponding author (email: [ning\\_ming@zju.edu.cn](mailto:ning_ming@zju.edu.cn))

to the same cluster have at least one common channel available.

(iii) Efficient source sensing: the sensed source information can be aggregated to the sink node.

(iv) Energy efficiency: the clusters are organized such that the total communication power is minimized.

To accommodate the unique objectives and requirements of the CRSN, we propose an energy-efficient spectrum-aware clustering scheme by jointly considering its source-sensing task and spectrum-aware nature. The network data communication can be categorized into two classes: intra-cluster communication and intercluster communication. During the intracluster communication phase, all the CRSN nodes send their readings of source information to their cluster head (CH) through the local common channel. During the intercluster communication phase, the CH transmits the aggregated information to the upstream neighbor CH using higher power.

## 1 Network energy-consumption model

The network energy consumption comprises three parts: multi-task sensing, signaling and data transmission. We assume that there are  $N$  CRSN nodes and  $K$  clusters. The  $k$ th cluster is denoted  $c_k$  and has  $N_k$  CRSN nodes. The  $i$ th node of  $c_k$  is  $n_i^k$ , whose coordinates are  $(x_i, y_i)$ .

The multi-task sensing-power optimization problem has already been investigated [8] and is independent of the clustering scheme. For ease of analysis, we assume that the sensing result at each node is always correct and we denote the minimal multi-task sensing power as  $P_{MS}$ .

The signaling power is mainly used to convey the control information, which is crucial for the cluster formation, CH selection and channel/distance estimation. These beacon signals are broadcast at fixed maximal power in available channels. The total power is denoted  $P_{SG}$ .

The data transmission comprises intracluster communication and intercluster communication. In the intracluster communication stage, all CRSN nodes send their readings to the CHs through common channels of the cluster. When the  $j$ th node is selected as the CH (i.e.  $i_k=j$ ), the intracluster communication energy of the  $k$ th cluster is the sum transmission power of all cluster members (CMs):

$$P_{\text{intra}}(i_k = j) = \sum_{i=1, i \neq j}^{N_k} P_i(n_i^k). \quad (1)$$

In our scenario, all CMs adjust their transmission power properly to meet the required signal-to-noise ratio while saving energy. This can be achieved by first estimating the CM-to-CH channel fading and then finely adjusting the transmission power amplifier to a required level. Thus, the channel fading can be estimated by measuring the strength of the received "node beacon", while the required transmission power can be calculated using an appropriate channel

model. Since the distances between CMs and their CH are short within a cluster, a free-space channel model is applied with  $d^2$  power loss. The channel between a CM and CH is also assumed to be symmetric. Hence, the required transmission power for the  $i$ th node of  $c_k$  can be expressed as

$$P_i(n_i^k) = C_0 P_r d^2(n_i^k, n_j^k), \quad (2)$$

where  $P_r$  is the minimal received power required for CH to correctly decode the transmitted information;  $C_0$  is a constant loss factor and  $d(n_i^k, n_j^k)$  is the Euclidean distance between the  $i$ th and  $j$ th nodes.

When the  $j$ th node is selected as the CH, the intracluster power can be rewritten as

$$P_{\text{intra}}(\text{CH} = n_j^k) = \sum_{i=1, i \neq j}^{N_k} P_i(n_i^k) = C_0 P_r \sum_{i=1, i \neq j}^{N_k} d^2(n_i^k, n_j^k). \quad (3)$$

To balance the power consumption, the CRSN nodes take turns to act as the CH, and the probability of the  $j$ th node being the CH is denoted  $p(\text{CH} = n_j^k)$ . Statistically, the overall average intracluster power consumption of  $c_k$  is

$$\begin{aligned} P_{\text{intra}}(c_k) &= \sum_{j=1}^{N_k} p(\text{CH} = n_j^k) P_{\text{intra}}(\text{CH} = n_j^k) \\ &= C_0 P_r \sum_{j=1}^{N_k} p(\text{CH} = n_j^k) \sum_{i=1, i \neq j}^{N_k} d^2(n_i^k, n_j^k). \end{aligned} \quad (4)$$

After the CH has collected all sensed source information, it compresses and forwards the source information to the sink node through the vacant channels shared with upstream clusters in the intercluster communication stage. The intercluster communication power is fixed at maximal power to improve intercluster connectivity. The sum power for intercluster communication can be expressed as

$$P_{\text{inter}} = \sum_{k=1}^K P_{\text{IC}} = K C_0 P_r d_{\text{max}}^2, \quad (5)$$

where  $d_{\text{max}}$  is the transmission range of the CRSN node.

CH selection. Before each round of intracluster communication, we select the CRSN node with maximal residual energy as the CH. This maximal residual energy (MRE) rule aims to balance the power consumption among different CRSN nodes and can substantially extend the network lifespan.

Proposition 1. If the CRSN nodes are initialized with equal energy, and we select the CH according to the MRE rule, the average intracluster power can be upper-bounded by the inequality

$$P_{\text{intra}}(c_k) \leq C_0 P_r \sum_{j=1}^{N_k} \frac{1}{N_k} \sum_{i=1, i \neq j}^{N_k} d^2(n_i^k, n_j^k). \quad (6)$$

The proof is intuitive. First, it is obvious that the upper-bound can be achieved by fixing each CRSN node's proba-

bility of becoming the CH as  $1/N_k$ . We then show that the energy consumption can be further reduced when applying the MRE rule. As illustrated in Figure 1(a), we assume  $n_i$  is a near-center node and  $n_j$  is a near-border node. Although  $n_i$  and  $n_j$  consume the same maximal power in the intercluster stage when they are selected as the CH, the near-border  $n_j$  will consume more energy than  $n_i$  in the intracluster stage. This is because the average distance from  $n_j$  to the remaining nodes is longer than that for  $n_i$ . As this continues,  $n_j$  will consume more power and have less residual energy than  $n_i$ . According to the MRE rule,  $n_i$  will have a higher chance of being selected as the CH. When  $n_i$  becomes the CH, the sum CM-CH power for the remaining nodes is lower than when  $n_j$  is selected as the CH. Therefore, by selecting  $n_i$  nodes as the CH more often, we can further reduce the sum intra-cluster power. Moreover, the energy consumption is balanced among all CRSN nodes and the average lifetime is extended.

However, when all CRSN nodes happen to be symmetric, as shown in Figure 1(b), each node will have an equal probability of becoming CH and the intracluster energy thus achieves the upper-bound even if the MRE rule is employed. Since the sensor nodes are usually randomly deployed and their structure is unknown beforehand, we define the sum intracluster power using its upper bound. After some further derivation, the sum intracluster power consumption can be simplified as

$$P_{\text{intra}} \triangleq \sum_{k=1}^K C_0 P_r \sum_{j=1}^{N_k} \frac{1}{N_k} \sum_{i=1, i \neq j}^{N_k} d^2(n_i^k, n_j^k) = 2C_0 P_r \sum_{k=1}^K \sum_{i=1}^{N_k} d^2(n_i^k, \text{center}(k)), \quad (7)$$

where  $\text{center}(k) = \left( \frac{1}{N_k} \sum_{i=1}^{N_k} x_i^k, \frac{1}{N_k} \sum_{i=1}^{N_k} y_i^k \right)$  is the center of the  $k$ th cluster. Therefore, the network-wide energy consumption can be rewritten as

$$E_{\text{network}} = P_{\text{MS}} t_{\text{MS}} + P_{\text{SG}} t_{\text{SG}} + \left( KC_0 P_r d_{\text{max}}^2 + 2C_0 P_r \sum_{k=1}^K \sum_{j=1}^{N_k} d^2(n_i^k, \text{center}(k)) \right) t_{\text{DATA}}, \quad (8)$$

where  $t_{\text{MS}}$  is the sensing time,  $t_{\text{SG}}$  is the signaling time, and  $t_{\text{DATA}}$  is the data transmission time.

Since the multi-task sensing power and signaling power do not depend on the choice of clustering scheme, the average network-wide energy is mainly affected by the data transmission power from the clustering perspective; i.e.  $\text{Minimize}(E_{\text{network}}) \Leftrightarrow \text{Minimize}(P_{\text{inter}} + P_{\text{intra}})$ .

## 2 Energy-efficient optimization problem

### 2.1 Optimal number of clusters

We have  $N$  CRSN nodes and need to partition them into  $K$  clusters. How many clusters should be created is a critical consideration in our energy-saving task. For instance, if  $K=N$ , then each CRSN node is an independent cluster and all CRSN nodes act as CHs and have to transmit at maximal power. By contrast, if  $K=1$ , then all  $N$  CRSN nodes form a single cluster and the intracluster communication energy will be too high owing to long intracluster distances. Both extreme cases will result in excessive energy consumption. Therefore, the optimal number of clusters should be properly chosen to minimize network-wide energy. For uniformly distributed CRSN nodes, we can analytically derive the optimal number of clusters.

The expectation of data transmission power is

$$E(P_{\text{DATA}}) = 2C_0 P_r E \left( \sum_{k=1}^K \sum_{i=1}^{N_k} d^2(n_i^k, \text{center}(k)) \right) + KC_0 P_r d_{\text{max}}^2. \quad (9)$$

It is reasonable to assume that the randomly deployed CRSN nodes are uniformly distributed in the two-dimensional area around the center point, and the density  $\rho$  is predetermined by the application-specific source-sensing requirement. Therefore

$$E \left( \sum_{k=1}^K \sum_{i=1}^{N_k} d^2(n_i^k, \text{center}(k)) \right) = N(\text{var}(x_i^k) + \text{var}(y_i^k)) = \frac{Nd^2}{6}, \quad (10)$$

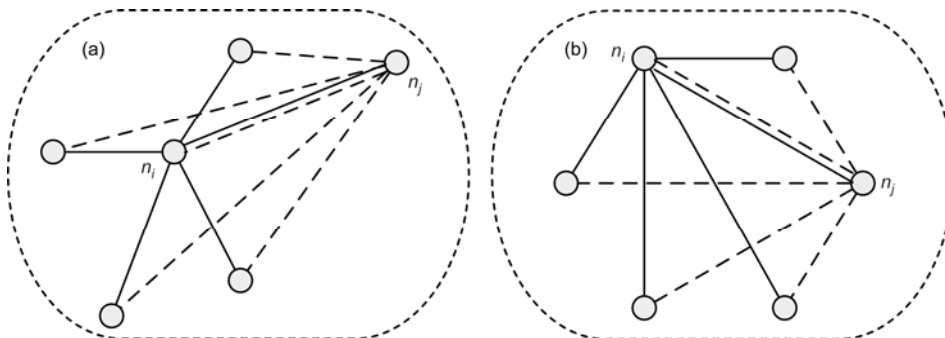


Figure 1 Intracluster communication. (a) Asymmetry cluster; (b) symmetry cluster.

where  $d$  is the average span of a cluster. Since there are  $N/K$  nodes per cluster on average, and the density of CRSN nodes is  $\rho$ , then the area of the cluster can be estimated as  $d^2 = N/K\rho$ .

Substituting the above formulations into eq. (9), we get

$$E(P_{\text{DATA}}) = C_0 P_r \left( \frac{N^2}{3\rho K} + Kd_{\text{max}}^2 \right). \quad (11)$$

Thus, the optimal number of clusters can be estimated by setting the derivative of eq. (11) with respect to  $K$  to zero. The result should be rounded to an integer:

$$K_{\text{opt}} = \left\lfloor \frac{\sqrt{N}}{\sqrt{3}d_{\text{max}}} + 0.5 \right\rfloor. \quad (12)$$

## 2.2 Clustering to minimize the network-wide energy

After the optimal number of clusters  $K_{\text{opt}}$  is determined, we find that the network-wide energy is mainly affected by intracluster communication energy. Hence, according to eq. (7), minimizing the network-wide energy is equivalent to minimizing the sum of the squared distance between CRSN nodes and their cluster centers:

$$\text{Minimize } E_{\text{network}} \Leftrightarrow \text{Minimize } \sum_{k=1}^K \sum_{i=1}^{N_k} d^2(n_i^k, \text{center}(k)). \quad (13)$$

In a CRSN, only the nodes having the common available channel and within the maximum transmission range  $d_{\text{max}}$  can form a cluster. These requirements are imposed on the optimal clustering problem as spectrum-aware constraints and transmission-range constraints. Therefore, the optimal clustering can be obtained from

$$\arg \min_{C \in \mathcal{C}} \sum_{k=1}^K \sum_{i=1}^{N_k} d^2(n_i^k, \text{center}(k)), \quad (14)$$

subject to

$$\text{C1: } |\text{Chan}(n_i^k) \cap \text{Chan}(n_j^k)| \geq 1, n_i^k, n_j^k \in \text{Cluster}(k);$$

$$\text{C2: } \max_{n_i^k, n_j^k \in \text{Cluster}(k)} d(n_i^k, n_j^k) < d_{\text{max}},$$

where  $|\text{Chan}(n_i^k)|$  denotes the number of available channels for  $n_i^k$ ,  $\mathcal{C}$  is the set of all possible clustering results and  $C$  represents a specific clustering in the set. The former constraint is a spectrum-aware constraint that ensures the availability of a common channel within a cluster, and the latter is the maximum-transmission-range constraint.

One way of finding the optimal solution is to try all possible ways of clustering options, and choose the one that minimizes the sum of the squared distance. This exhaustive search method can help us find the global minimum, but the complexity is too high. Supposing we are given  $N$  points and would like to partition them into  $K$  clusters, then there is a maximum of  $K^N/K!$  clustering options. As a result, an

exhaustive search would be computationally infeasible in practice, and we have to turn to other more efficient suboptimal clustering algorithms.

## 3 Spectrum-aware constrained clustering

In this section, we assume error-free spectrum sensing for the ease of analysis. The sensing results are imposed on secondary users as the spectrum-aware constraint, which is regarded as a groupwise constraint, and novel methods are proposed to solve the problem.

### 3.1 Groupwise constrained clustering

In clustering analysis theory, the objective function in eq. (14) is called the sum of the squared error (SSE), or ‘‘scatter’’. Minimizing the SSE is also the goal of many clustering algorithms. Therefore, we can employ concepts in clustering analysis theory to design desirable clustering schemes. Some computationally feasible heuristic methods have been developed for efficient clustering. The main techniques are  $k$ -means, fuzzy  $c$ -means, and hierarchical clustering, etc. Several of them are effective in clustering non-cognitive WSNs [7]. However, none of these algorithms consider the spectrum-aware constraint. In recent years, a branch of constrained clustering algorithms has been developed to cluster instances with pairwise constraints, such as constrained  $k$ -means [9] and constrained complete-link clustering [10]. The pairwise constraint is imposed on pairs of nodes to influence the outcome of the clustering algorithm, which mainly involves two types of constraints: must-link and cannot-link constraints.

As shown in Figure 2, the must-link constraint forces  $n_i$  and  $n_j$  to be in the same cluster and the cannot-link constraint specifies that  $n_i$  and  $n_j$  must not be placed in the same cluster. If two CRSN nodes have no available channels in common, they cannot be allocated into one cluster, and this is equivalent to imposing a cannot-link constraint on this node pair. Thus, the ideas of constrained clustering algorithms can be used to design a spectrum-aware clustering scheme for the CRSN. However, the existing constrained clustering methods cannot be directly applied to the spectrum-aware clustering, since our spectrum-aware constraints are imposed on groups, rather than on pairs.

We now define the ‘‘groupwise constraint’’ by explaining the differences between the ‘‘pairwise constraint’’ and ‘‘groupwise constraint’’. In Figure 2, three nodes can operate in three channels, and the numbers given beside the nodes represent the available channels. In the middle and on the right, nodes A and B share channel 1, nodes A and C share channel 2, and nodes B and C share channel 3. If employing a pairwise constraint, each node pair shares a common channel and no ‘‘cannot-link constraint’’ is imposed on them, and they can form one cluster. However, if a groupwise

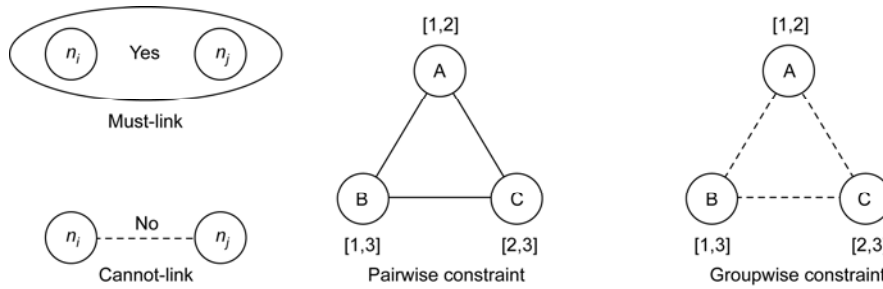


Figure 2 Comparison of the pairwise constraint and groupwise constraint.

constraint is imposed, the three nodes share no common channel and cannot form a cluster. To extend the existing algorithms to the model with a groupwise constraint, we have to replace the pairwise constraint with a spectrum-aware groupwise constraint.

### 3.2 Centralized spectrum-aware clustering algorithm

We first propose a centralized spectrum-aware clustering (CSAC) algorithm for the CRSN. The algorithm is inspired by the constrained complete-link agglomerative clustering algorithm [10], which treats each node as an individual cluster at the beginning and then merges the two nearest clusters in each iteration. However, we incorporate the spectrum-aware groupwise constraint instead of the pairwise constraint. The network topology is hierarchical as in many large-scale WSNs [5]. CRSN node information, which includes geographical position (e.g. that obtained through the global positioning system (GPS)), available channels and residual energy, is aggregated to a central processor such as the base station (BS), through the detected vacant channels. Afterward, the BS implements the centralized clustering algorithm and then informs each node of the clustering result.

The CSAC algorithm is described in Table 1. In the initializing phase, each node senses the available channel, and reports its own node information to the BS. The BS will create a proximity matrix according to the pairwise distance. During the iterations, the spectrum-aware constraint is incorporated by setting the proximity matrix based on the groupwise constraint: if the CRSN nodes in two clusters share no common channel, the corresponding distance entry in the proximity matrix is set to infinity; otherwise the complete-link rule [10] is implemented. The nearest clusters that satisfy the groupwise constraint are then merged. The two rows and columns representing these two clusters are merged as new ones in the proximity matrix. The iterative merging process stops when the cluster number reaches  $K_{opt}$  or no merge satisfying the constraints can be performed. In the end, the CH is selected according to the MRE rule, and the clustering result is broadcast to all CRSN nodes.

### 3.3 Distributed spectrum-aware clustering protocol

Although the centralized clustering algorithm can effectively

Table 1 CSAC algorithm

| Algorithm 1 Centralized spectrum-aware clustering |   |
|---|---|
| Step 1: Perform channel sensing                   |   |
|   | Chan( $n_i$ )=Sensed result.  |
| Step 2: Report node information                   |   |
|   | $n_i \leftarrow (x_i, y_i), i \in \{1, 2, \dots, N\}$ ;   |
| Step 3: Create proximity matrix                   |   |
|   | $D \leftarrow d_{i,j} =  X_i - X_j $ ;  |
| Step 4: Spectrum-aware constrained clustering     |   |
|   | a) Initialize each node as a cluster: $c_i(n_i) \leftarrow n_i$   |
|   | b) Intercluster distance: $d(c_m, c_n) = \max_{n_i \in c_m, n_j \in c_n} d(n_i, n_j)$ ;   |
|   | c) Impose groupwise spectrum-aware constraint   |
|   | for $i, j \in c_k$ do   |
|   | if $ \text{Chan}(n_i) \cap \text{Chan}(n_j)  = 0$ then  |
|   | $d_{i,j} \leftarrow \infty$ ;   |
|   | end   |
|   | end   |
|   | d) Find the nearest clusters: $(m, n) = \arg \min_{m, n \in \{1, 2, \dots, K\}} d(c_m, c_n)$  |
|   | e) Merge $c_m$ and $c_n$ into $c_l$ : delete the $m$ th and $n$ th rows and columns in $D$ ; add the $l$ th row and column for the newly formed cluster $c_l$ . |
|   | f) Update $D$ : $d(c_k, c_l) \leftarrow \max(d(c_k, c_m), d(c_k, c_n))$ ;   |
|   | g) Go to sub-step c).   |
| Step 5: Select cluster head                       |   |
|   | $\text{CH}_k = \arg \max_{i \in c_k} \{E_i\}$ .   |

partition the CRSN into clusters that satisfy the spectrum-aware constraint, it has two major drawbacks. First, massive signal exchange between the CH and BS is required to collect network-wide node information. In practice, it is difficult to find a wireless channel for direct connection between the CH and BS, since such a connection requires much higher transmission power and will bring interference to a nearby primary-user (PU) system. Moreover, the complexity of the CSAC algorithm is proportional to the size of the proximity matrix, which is  $N^2$ . As a result, the growing size of the network will result in greater complexity, making the

algorithm difficult to implement for large-scale networks. Obviously, these shortcomings will inevitably limit the practical application of this algorithm.

To address these problems, we propose a novel distributed spectrum-aware clustering (DSAC) technique, with even lower complexity and higher stability. The basic idea of DSAC remains the same as that of CSAC in general: the closest nodes with common channels will agglomerate into a small group first and then the other neighboring nodes will join in one after another. The main differences are that CSAC compares the distance between all clusters and finds the global minimum pair to merge first, while DSAC only needs to determine the local minimum distance through neighborhood information exchange and merges the locally closest pair.

The DSAC protocol is described by the flowchart in Figure 3. It consists of three stages: channel sensing, beaconing and coordination. In the channel-sensing stage, each CRSN node determines the vacant channels individually and compares it with the previously sensed result. In the beaconing stage, the CRSN node beacons its node information in the vacant channels according to the channel-sensing results. If any change in the PU state is detected, the node declares itself as a new cluster by beacons a new cluster ID. Otherwise, the node stays with the current cluster. After the node beaconing, the CH updates and beacons the cluster information, including cluster size and common channels. In the intracluster coordination stage, each node in a cluster first measures the strength of neighboring beacon signals and then announces the pairwise distances. From these distances, the CH determines the intercluster distance according to the complete-link rule and the groupwise constraint, in which the intercluster distance is defined as the maximum distance between the nodes of two clusters. In intercluster coordination, each CH sends merge invitation to the nearest neighboring cluster that is within its transmission range. If any two clusters send merge invitations to each other, they merge into a single cluster by unifying the new cluster ID and common channels, and selecting a new CH with the largest residual energy. Otherwise, the cluster selects a new CH while the topology remains unchanged.

Figure 4 shows an example of the DSAC clustering result, where 50 CRSN nodes and 10 PUs are randomly deployed on a 100 m×100 m field. There are three available channels in the system (colored red, green and blue). The clustering result is illustrated by a dashed circle and the corresponding common channels are labeled in the cluster.

### 4 Performance evaluation

#### 4.1 Complexity and scalability

The complexity of the CSAC algorithm is similar to that of existing complete-link agglomerative clustering algorithms. The time complexity of the CSAC algorithm is  $O(N^2 \log N)$

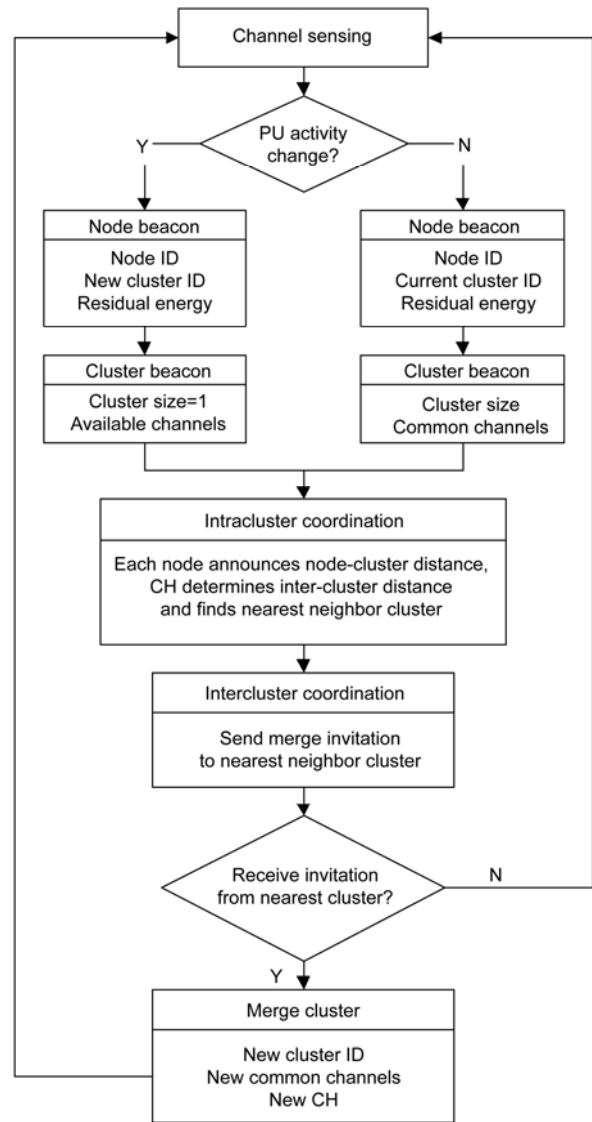


Figure 3 Flowchart of the DSAC protocol.

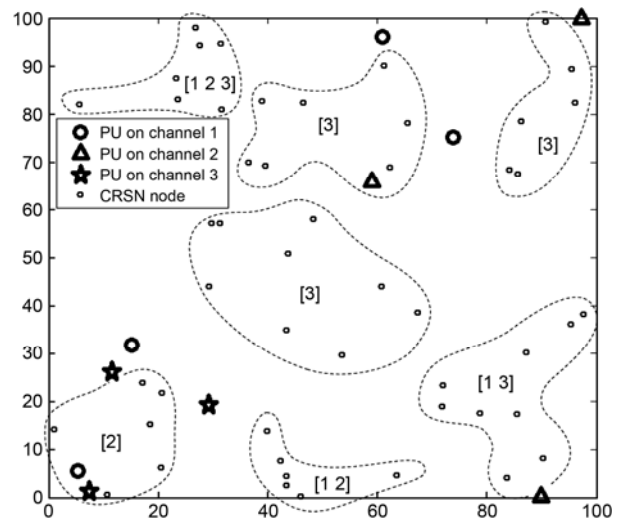


Figure 4 Example of the DSAC clustering result.

and the space complexity is  $O(N^2)$ . This is because the size of proximity matrix grows with the square of the number of CRSN nodes. Although this complexity is much lower than that of the optimal algorithm and the algorithm can be well implemented for some small sensor networks, the complexity is still too high for the algorithm to be implemented for a large-scale CRSN.

The DSAC protocol is a variation of the CSAC algorithm. On the one hand, while CSAC merges the global minimum distance pair from the  $N \times N$  proximity matrix, DSAC only seeks to find the local nearest cluster pair from the neighboring clusters. Therefore, its space complexity is the number of neighboring clusters and it will not grow rapidly with the size of the whole network. On the other hand, CSAC only merges one cluster pair in each iteration as the minimum distance is unique within the network. However, in DSAC, more than one of the local nearest cluster pairs can be found and multiple merges may occur in one cycle, which also increases the convergence rate of the DSAC protocol. As a result, the time complexity of the DSAC protocol is much less than that of the CSAC algorithm.

Figure 5 demonstrates the convergence of DSAC and CSAC schemes. The CRSN size means the number of CRSN nodes. At the beginning, every node is treated as a cluster, and during the following iterations, they gradually merge to form required clusterings. We find that DSAC converges to the desired number of clusters much more quickly than CSAC.

Figure 6 simulates the average converge time of the three clustering schemes as the CRSN size increases. The converging time of CSAC grows proportionally with the CRSN size, while DSAC converges almost as quickly as the efficient  $k$ -means algorithm. This result also shows the satisfactory scalability of DSAC.

**4.2 Energy consumption**

In the simulation, we use the sub-optimal  $k$ -means clustering

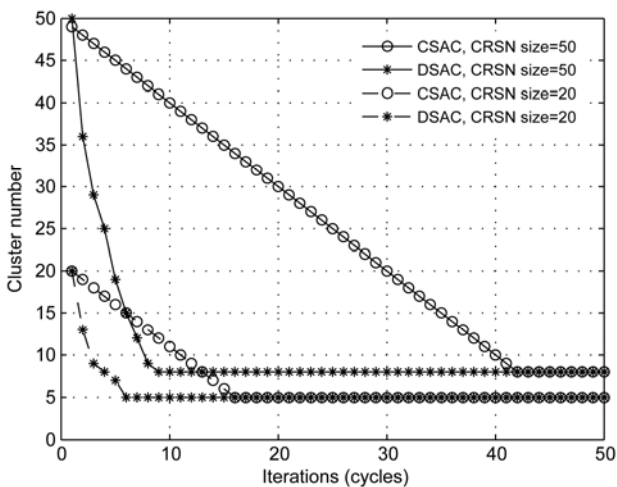


Figure 5 Time complexity: iterations vs. cluster number.

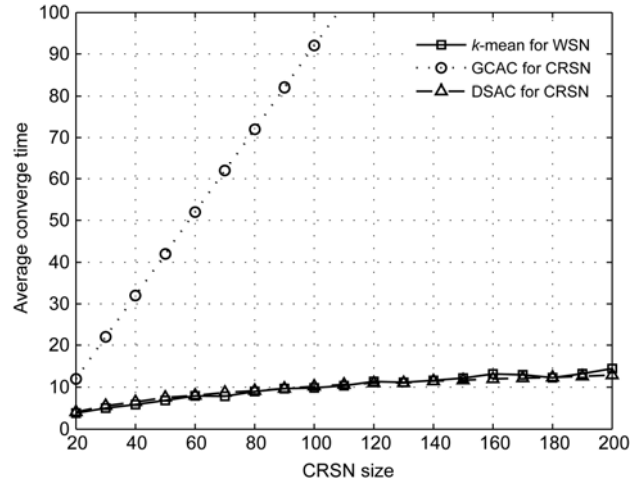


Figure 6 Scalability: CRSN size vs. converge time.

algorithm as the reference, since it converges to local minimal SSE in short time. Although  $k$ -means clustering does not consider the spectrum-aware constraint in the CRSN and is only applicable for non-cognitive WSNs, it serves as a universally accepted reference for the performance evaluation of our proposed clustering schemes.

In all the experiments, we randomly deploy PUs and CRSN nodes in a 100 m  $\times$  100 m area. The PUs can operate in three channels, and CRSN nodes can only access the channels in which the neighboring PUs are inactive. Every PU randomly occupies one of the three channels. The protection range for PU is 20 m, which means the PU's CRSN neighbors within this range cannot access its occupied channel.

As explained before, as long as the cluster number is determined, the performance in terms of energy efficiency is mainly decided by the SSE of the clustering result. The clustering with less SSE is more energy efficient. In Figure 7, we assume there are five PU nodes and six clusters and we compare the average SSE with respect to different CRSN sizes of the three clustering schemes. The experiment demonstrates the effectiveness of our two spectrum-aware clustering schemes by comparing with the unconstrained  $k$ -means clustering scheme as a reference. We find that both CSAC and DSAC schemes for the CRSN have clustering performance comparable to that of  $k$ -means for non-cognitive WSNs in terms of SSE. Moreover, the gap between DSAC and CSAC is rather narrow, which means the distributed scheme is well designed.

In the following experiments, we assume the maximum transmission range for a CRSN node is 50 m, and 20 CRSN nodes and five PU nodes are uniformly distributed in the 100 m  $\times$  100 m area. According to the previous theoretical analysis, the estimated optimal cluster number is about five. In the simulation, we set the cluster number from three to eight, and calculate the average power consumed by CRSN nodes. From Figure 8, we find that the power is a minimum

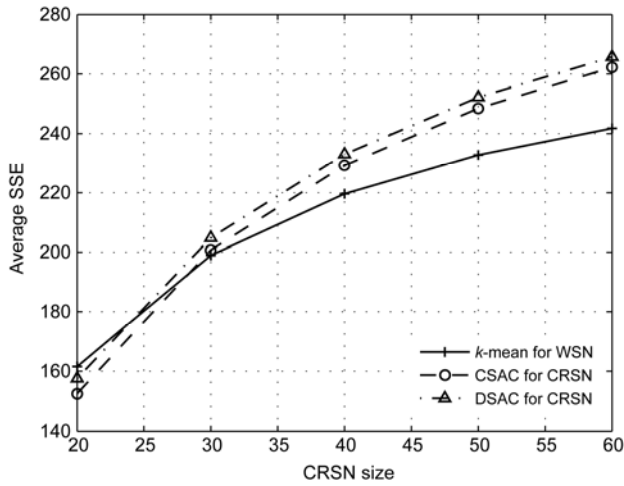


Figure 7 CRSN size vs. SSE.

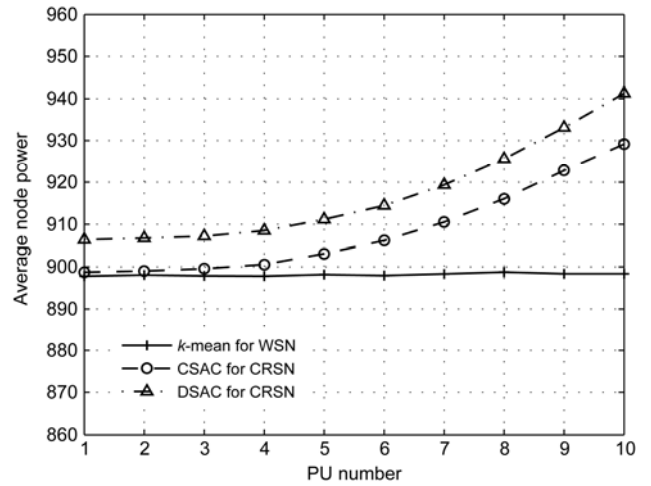


Figure 9 PU number vs. average node power.

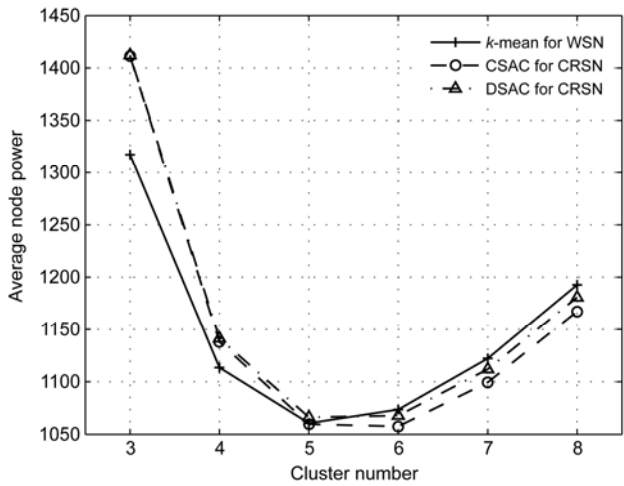


Figure 8 Comparison of average energy among three schemes.

when the cluster number is about five or six, and this simulation result agrees well with the analysis.

To evaluate the effect of PU activities on the CRSN, we simulate the average CRSN node power consumption when different numbers of PU nodes are active. In Figure 9, we set the CRSN size as 30 and adjust the PU number from 1 to 10. For a non-cognitive WSN, the *k*-means clustering result is not affected by the presence of PU systems, and the average node power consumption thus remains steady. For the CRSN, as more PU nodes become active, more spectrum-aware constraints are imposed on the clustering process. Therefore, the clustering results are poorer in terms of energy consumption. Again, we find the performance of DSAC only to be slightly worse than that of CSAC.

### 4.3 Stability against dynamic PU activities

Although the PU activity pattern usually remains steady within a very short interval, it will eventually vary with time.

Moreover, the CRSN node mobility may also contribute to the dynamic ambient radio environment. The dynamic PU activity and CRSN mobility will cause unstable clustering topology and frequent network failure, which induce excessive control overheads and poor connectivity. Fortunately, we find that the DSAC protocol can minimize this effect and maintain the network stability as much as possible.

For the optimal clustering algorithm, *k*-means clustering and CSAC, if any PU activity or CRSN node position changes, the whole network should be involved in re-clustering, which makes the network topology less stable and requires extra control overheads. However, according to the DSAC protocol, only the nodes that detect PU activity change will engage in re-clustering, and all other CRSN nodes and their topology remains stable. In our experiment, when one PU changes status, only 2 of the 50 CRSN nodes are affected, as shown in Figure 10. After two merges, the network again converges to stable clustering, which is much faster than the case for CSAC. During the re-clustering,

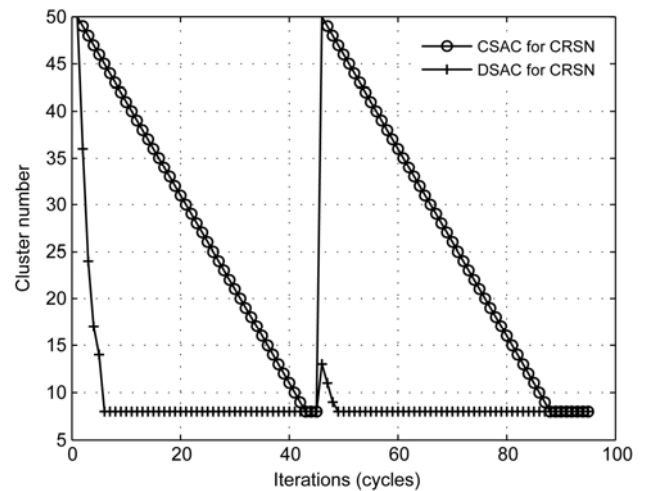


Figure 10 Comparison of dynamic stability between CSAC and DSAC.



the statuses of the remaining nodes and their clustering topology remain the same. Their application-specific sensing task is not influenced. The stability of the network is thus preserved as much as possible.

## 5 Conclusions

In this paper, we proposed two novel spectrum-aware clustering schemes for the CRSNs [11]. We first modeled the energy consumption particularly for the CRSN. After deriving the optimal number of clusters, we found that the CRSN energy minimization problem can be converted to an SSE minimization problem. By employing the groupwise constraints to characterize the spectrum-aware requirements, we proposed the CSAC algorithm and the DSAC protocol to solve the minimization problem. Through extensive simulation results, we further verified the effectiveness and efficiency of our approach.

In the future works, theoretical performance analysis of the algorithm is desired. The adaptations of various PU spectrum occupation models can be investigated to acquire more robust performance. Moreover, experiments on testbeds are necessary before any practical implementations can be carried out in the future. In summary, this paper addresses the problem of distributed clustering in a dynamic wireless environment, which sheds light on the study of future cognitive radio systems.

*This work was supported by the National Basic Research Program of China (2009CB320405 and 2012CB316104), the National Natural Science Foundation of China (60972057), and the Supporting Program for New Century Excellent Talents in Universities (NCET-09-0701).*

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

- 1 Tian J F, Zheng X Y, Hu H L, et al. A survey of next generation mobile communications research in China. *Chin Sci Bull*, 2011, 56: 2875–2888
- 2 Akan O, Karli O, Ergul O, et al. Cognitive radio sensor networks. *IEEE Network*, 2009, 23: 34–40
- 3 Baddour K E, Ureten O, Willink T J. Efficient clustering of cognitive radio networks using affinity propagation. In: *Proceedings of the 18th International Conference on Computer Communications and Networks*, 2009 Aug 2–6, San Francisco. Washington DC: IEEE, 2009. 1–6
- 4 Zhang J Z, Wang F, Yao F Q, et al. Cluster-based distributed topology management in cognitive radio ad hoc networks. In: *Proceedings of International Conference on Computer Application and System Modeling*, 2010 Oct 22–24, Taiyuan. New York: ASME Press, 2010. 544–548
- 5 Heinzelman W B, Chandrakasan A P, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans Wirel Commun*, 2009, 1: 660–670
- 6 Younis O, Fahmy S. HEED a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans Mob Comput*, 2004, 3: 366–379
- 7 Gong Y L, Chen G, Tan L S. A balanced serial k-means based clustering protocol for wireless sensor networks. In: *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 2008 Oct 12–14, Dalian. Washington DC: IEEE, 2008. 1–6
- 8 Zhang H Z, Zhang Z Y, Chen X M, et al. Energy efficient joint source and channel sensing in cognitive radio sensor networks. In: *Proceedings of International Conference on Communications*, 2011 June 5–9, Kyoto. Washington DC: IEEE, 2011. 1–6
- 9 Wagstaff K, Cardie C, Rogers S, et al. Constrained *k*-means clustering with background knowledge. In: *Proceedings of the 18th International Conference on Machine Learning*, 2001 June 28–30, Williamstown. San Francisco: Morgan Kaufmann Publishers, 2001. 577–584
- 10 Klein D, Kamvar S D, Manning C D. From instance-level constraints to space-level constraints: Making the most of prior. In: *Proceedings of the 19th International Conference on Machine Learning*, 2002 July 8–12, Sydney. San Francisco: Morgan Kaufmann Publishers, 2002. 307–314
- 11 Zhang H Z, Zhang Z Y, Dai H Y, et al. Distributed spectrum-aware clustering in cognitive radio sensor networks. In: *Proceedings of IEEE Globecom 2011 Conference*, 2011 Dec 5–9, Houston. Washington DC: IEEE, 2011. 1–6