

Received June 29, 2019, accepted July 15, 2019, date of publication July 22, 2019, date of current version August 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929574

A Spectrum-Aware Clustering Algorithm Based on Weighted Clustering Metric in Cognitive Radio Sensor Networks

TIANJING WANG¹, XINJIE GUAN¹, XILI WAN¹, HANG SHEN¹, AND XIAOMEI ZHU

School of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China

Corresponding author: Tianjing Wang (wangtianjing@njtech.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501224, Grant 61802176, Grant 61602235, Grant 61501223, and Grant 61502230, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20161007 and Grant BK20150960.

ABSTRACT Clustering organizes nodes into groups in order to enhance the connectivity and stability of cognitive radio sensor networks. Mainly depending on the channel availability, many existing spectrum-aware clustering algorithms may not achieve the most satisfactory clustering. Taking into account the various influence factors to establish the optimal clustering is a challenge to enhance the network performance. This paper proposes a novel spectrum-aware clustering algorithm based on weighted clustering metric to obtain the optimal clustering by solving an optimization model. The new weighted clustering metric, simultaneously evaluating temporal-spatial correlation, confidence level and residual energy, is used to elect clusterheads and ally member nodes. After clustering, the clusterheads sensing spectrum instead of all member nodes greatly reduces the energy consumption of spectrum sensing and increases the opportunity of data transmission. The performance comparison between the traditional spectrum-aware clustering algorithms and our proposed algorithm has been highlighted with the experiments.

INDEX TERMS Cognitive radio sensor networks, spectrum-aware clustering, weighted clustering metric, temporal-spatial correlation, confidence level.

I. INTRODUCTION

Tremendous growths of technologies and applications for wireless sensor networks (WSNs) have brought significantly increased demand for radio spectrum, but the traditional static spectrum allocation policies have led to the spectrum scarcity [1]–[3]. The cognitive radio sensor network (CRSN) is an emergent paradigm that efficiently alleviates spectrum scarcity and better utilizes available spectrum [4]. Equipped with the cognitive radio module, the cognitive sensor nodes (CSNs) can opportunistically access the spectrum bands licensed to the primary users (PUs) without interfering with them.

It has known that the CRSN inherits some intrinsic properties of WSNs (e.g., the limitation of hardware and energy resource). Therefore, it is difficult to maintain the connectivity and stability of CRSN when the energy consumption for transmitting a large amount of mutual information and

application data leads to the invalidation of some nodes [5]. To reduce the resource overheads of spectrum sensing, routing and data transmission, clustering is widely applied to scale down network and enhance the network performance. We observe that clustering can bring about three main advantages to CRSNs, namely scalability, stability and supporting cooperative tasks [6], [7]. However, the traditional clustering algorithms (e.g., lowest ID [8], maximum node degree [9], and K -mean clustering [10]) cannot be directly applied to CRSNs because of operating on a fixed channel. The Low Energy Adaptive Hierarchy (LEACH) was a base line distributed clustering scheme for WSNs [11], [12]. Dealing with the dynamic channel change of CRSNs, a spectrum-aware clustering protocol termed as CogLEACH was designed based on LEACH [13]. A low-energy adaptive uneven clustering hierarchy (LEAUCH) considered the advantage of the available channel resource to reduce the energy consumption [14]. In [15], a RARE protocol split the network into clusters by defining cluster formation as a maximum edge biclique problem. According to spectrum

The associate editor coordinating the review of this manuscript and approving it for publication was Kaigui Bian.

availability and nodes mobility, clusters in RARE could maintain the integrity of the network framework. Compared to clustering in WSNs, it is necessary to apply the channel availability to perform cluster formation and cluster maintenance in CRSNs.

A. PROBLEM STATEMENT

Recently, many spectrum-aware clustering algorithms have been proposed to tackle the clustering problem in CRSNs. In [16], the nodes were grouped into clusters based on a highly connected subgraph where the value of each edge was defined as the number of the common idle channels between two nodes. In [17], the nodes exchanged information, including IDs and available channel lists, with their one-hop and two-hop neighbours to form local clusters, and then a spectrum-aware routing was constructed by comparing the available channel lists of adjacent nodes. In [18], a novel non-parametric Bayesian-based channel clustering scheme was proposed to identify the supported QoS levels over multiple available licensed channels, and thus SUs could identify an appropriate cluster or set of licensed channels that fulfilled their stringent QoS requirements. In [19], SMART triggered re-clustering by merging or splitting some of clusters due to the change of available channels. In [20], all second users (SUs) sensed the available channels and gathered the information of neighbour nodes using HELLO packets. SUs checked their own maintained information set S to perform clusterhead election and node join operation. Simply using the channel availability without analysing the similarity property of the neighboring nodes' sensing results, the above spectrum-aware clustering algorithms may provide sub-optimal clustering. In [21], a ROSS clustering algorithm was proposed on the basis of the proximity of available channels between the SUs and their neighbours, where the individual and neighboring connectivity degrees were used to elect clusterheads. However, the proximity of available channels was defined only comparing the sensed idle channels of a SU with that of its neighbours, which could not completely illuminate the similarity property of spectrum sensing in intra-cluster. **Thus, establishing more reasonable similarity model is an essential problem to construct satisfactory clustering.**

To solve this problem, a spatial similarity model was provided to approximate the number of idle channels available to each node. The nodes with high spatial similarity thus formed a cluster [13]. In [22], some physically close nodes could sense more same available channels, so they were grouped together by measuring the clustering metric, i.e., their Euclidean distances. Considering the available channels, geographical positions and database statistics, an improved spectrum-aware clustering algorithm for centralized cognitive radio networks (CRNs) was proposed to improve the network throughput and maintain the stability of clusters [23]. On the other hand, a temporal similarity model was illustrated by a bipartite graph presented the essential property of spectrum-aware clustering [24]. When the idle channels in two sets of vertices were highly temporal

correlation, the edges of these vertices could connect to form a cluster. As discussed, utilizing the spatial or temporal similarity model to form clusters is benefit to improve the network performance. Separately utilizing the spatial correlation and temporal correlation, however, these clustering methods do not always provide the most reasonable spectrum-aware clustering. Authors of [25] illuminated that sensor networks utilized clustering to aggregate gathered information at the cluster heads by exploiting correlation. Hence, clustering can exploit spatial and temporal correlation of sensor nodes' observations on spectrum sensing. In [26], a distributed spectrum-aware clustering (DSAC) protocol formed group-wise constrained clusters, where the optimal number of clusters was derived to minimize the intra-cluster distances under the spectrum-aware constraint. In [27], the temporal-spatial correlation of sensing results was transformed into the intra-cluster and inter-cluster sensing similarity such that an optimization problem with respect to the similarity metric was presented to determine the optimal number of clusters. We see that the temporal-spatial correlation is a very noteworthy characteristic of spectrum-aware clustering, but few existing works perform spectrum-aware clustering based on temporal-spatial correlation. **It is required to establish a clustering metric by modelling the temporal-spatial correlation of spectrum sensing.**

The above frequently-used clustering algorithms can provide spectrum-aware clustering for CRSNs without unreliable nodes, but the established clustering structure will be broken down using the sensing results of unreliable nodes. Specifically, some malicious nodes send the tampered sensing results to attack the spectrum decisions of normal nodes, and some nodes affected by the multi-path or shadow fading problem sending the biased sensing results will disturb the spectrum decisions of their neighbours. To exclude these unreliable nodes, some researchers have defined a clustering metric with respect to the reliability of nodes as the clustering criterion. In [28], a trust aware model involving sensing reliability and energy reliability was used to clustering trustworthy nodes and elect the most credible nodes as CHs. Thus, the model could reduce the handoff rate and improve the network throughput. In [29], the reliability ratio between the number of correct decisions and the total number of taken decisions was defined as a clustering metric to elect CHs. An update formula for the reliability parameter was introduced in [27]. When the value of reliability parameter exceeded a given threshold, re-clustering was triggered to delete the unreliable nodes. However, the frequent re-clustering results in the instability of topological structure and much resource consumption. It is observed that reinforcing the effect of reliable nodes on spectrum-aware clustering can avoid frequent re-clustering. **To quickly recognize and exclude unreliable nodes, establishing a clustering metric by modelling the sensing reliability is an important problem.**

In general, most of CSNs are battery powered which is difficult to recharge or replace for some environmental reasons, so they should minimize the energy consumption of spectrum

sensing operations and normal activities in order to maximize their lifetime. An algorithm “Quadrant Based Energy Efficient Clustering Hierarchy” (QBEECH) was proposed to explain that energy efficient clustering could increase both the stability and the lifetime of the network much further [30]. In [31], an energy-efficient spectrum-aware clustering algorithm was proposed to achieve clustering by allowing each member node to learn the cooperative sensing costs of its neighboring nodes. It then gave an optimal clustering that minimized the network energy consumption. Based on CogLEACH algorithm, a centralized probabilistic clustering algorithm (CogLEACH-C) was introduced, where cluster-heads election depended not only on the number of sensed idle channels but also on the level of each node’s residual energy [32]. In [33], a distributed spectrum-aware clustering algorithm determined the optimal number of clusters by establishing a network-wide energy consumption model with respect to the residual energy. In [34], an energy-aware cluster-based routing protocol for CRSNs selected the gateway nodes with more energy and picked CHs due to the residual energy and available channels. Thus, the event data are routed through an energy-efficient and stable path from a source node to sink. In accord with some traditional energy-efficient clustering algorithms, the above spectrum-aware clustering algorithms have offered the same idea that the clustering metric need to involve the residual energy of nodes. In established clusters, the member nodes sensing spectrum quickly consumes their residual energy. *According to the characteristic of cooperative spectrum sensing, how to reduce the energy consumption as far as possible is worth more deep exploration.*

As aforementioned, to provide the optimal clustering, the clustering criterion needs to consider many influence factors rather than only the channel availability. A weight-based localized clustering algorithm partitioned the network into single-hop clusters on the basis of the channel availability, the SUs’ speed and the PUs’ interference levels [35]. In [36], a weighted sum of capacity rank, stability rank and quality rank values termed as reserved value (RV) was the final value of each link. A cluster-based MAC protocol used RV to form clusters. A network stability-aware clustering (NSAC) defined the weight by considering both spectrum dynamic and residual energy. One node with the largest weight was selected as cluster head, then this CH and its neighbors formed one cluster marked “clustered”. The remaining nodes updated their weights and repeated the above clustering process until all nodes had been marked “clustered” [37]. This sequential clustering, however, increases the delay of clustering. In this paper, we propose a new spectrum-aware clustering algorithm based on weighted clustering metric involving temporal-spatial correlation, confidence level as well as residual energy. In a distributed mode, the node with the larger clustering metric than its neighbors is elected as a qualified CH, and thus the optimal clustering is established by maximizing the intra-cluster similarity and the inter-cluster dissimilarity. After clustering, the CHs sense spectrum instead

of their member nodes and share the sensing information with them, while the member nodes do not need to detect spectrum all the time that greatly reduces the energy consumption of spectrum sensing and increases the time of data transmission.

Our **main contributions** of this paper are summarized as follows:

- Different from the traditional spectrum-aware clustering criterion mainly involving the channel availability, we introduce a novel weighted clustering metric taking into account three important influence factors, i.e., temporal-spatial correlation, sensing confidence and residual energy. Due to the weighted clustering metric, we can provide the optimal clustering according to the different application scenarios.
- An optimization model measuring the intra-cluster similarity and the inter-cluster dissimilarity is established to solve the optimal number of clusters. Our clustering hierarchy based on the optimal cluster size can provide better network performance than some traditional spectrum-aware clustering methods relying on the channel availability.
- After clustering, the CHs periodically sense spectrum instead of their member nodes. Thus, the member nodes can choose provisional dormancy to avoid rapidly running out of the residual energy and becoming invalidation. This cooperative spectrum sensing mode guarantees the stability of the network topology. On the other hand, the member nodes can also choose to transmit data in order to enhance the network throughput.

B. ORGANIZATION

The rest of this paper is organized as follows. Section 2 introduces four evaluation criteria to propose a new weighted clustering metric for spectrum-aware clustering. Section 3 presents a clusterhead election method to elect high-performance CHs and establishes an optimization model to solve the optimal number of clusters. In Section 4, the experiments are provided to evaluate the performance of our proposed algorithm, showing the available throughout gain.

II. SYSTEM MODEL

A. SENSING STRUCTURE FOR SPECTRUM-AWARE CLUSTERING

In general, a CRSN consists of a Base Station (BS) and some randomly distributed sensor nodes equipped with CR module. The spectrum resource in this area is divided into N_c channels sharing by N_p PUs and N_s sensor nodes, where these nodes are equipped with the computation module to process complicated mathematical operation. All nodes use the energy detection method [38] to detect the idle channels, and a clustering metric is applied to organize these nodes into K clusters. Because spectrum sensing causes additional energy consumption, Figure 1 proposes a new sensing structure for spectrum-aware clustering, in which the CHs periodically sense spectrum instead of all member nodes after clustering. More specially, once some of CHs find that the number of the

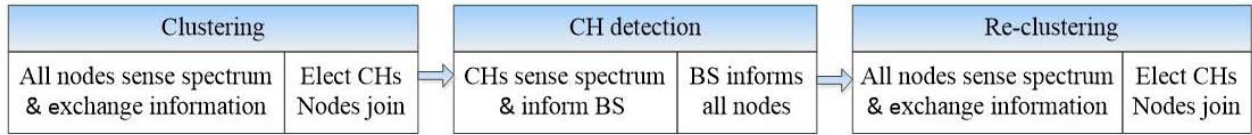


FIGURE 1. New sensing structure for spectrum-aware clustering in CRSNs.

common idle channels in their clusters is less than a threshold value, they immediately informs BS to broadcast the alert information about the change of channels. Receiving the message from BS, other CHs sense spectrum again such that the new spectrum sensing results trigger local re-clustering. This means that a CH make its member nodes join re-clustering if it also find the number of idle channels in its cluster reduces to the threshold, while a CH maintains its cluster if there is no change of idle channels in its cluster, and it only allows nodes outside cluster to join. In the other case, the BS will trigger the global re-clustering especially when the states of all PUs in the monitoring area change within a short time. Because the CHs have better sensing accuracy and more residual energy, this new sensing structure for spectrum-aware clustering can significantly reduce the network energy consumption without affecting the precision of spectrum sensing.

B. NEW EVALUATION CRITERIA FOR CLUSTERING

To create clusters, the similarity measurement between the observations of all nodes is required. In most clustering methods, this is achieved by use of the clustering metric. Hence, the choice of an appropriate clustering metric has a great impact on the quality of clustering. Focusing on channel availability, geographical location and residual energy, we define a new clustering metric to perform more reasonable clustering. Our clustering metric is evaluated by the following criteria.

1) SPATIAL CORRELATION:

The channel availabilities of all nodes are subject to their geographical locations, i.e., some geographically close nodes have higher probability in a cluster [39]. To illustrate the relationship of channel availability and geographical location, we define the spatial correlation between node *i* and node *j* in the *t*-th round of clustering as follows:

$$SC_{i,j,t} = \begin{cases} 1 - \frac{d_{i,j,t}}{\min\{r_i, r_j\}}, & d_{i,j,t} < \min\{r_i, r_j\} \\ 0, & d_{i,j,t} \geq \min\{r_i, r_j\} \end{cases} \quad (1)$$

where $d_{i,j,t}$ is the Euclidean distance between two nodes, $d_{i,j,t} < \min\{r_i, r_j\}$ means two nodes are in each other's communication range, and r_i, r_j are the communication radii of two nodes, respectively. To evaluate the mean spatial correlation between node *i* and its one-hop neighbours,

the expectation of spatial correlation is defined as

$$E(SC_{i,t}) = \frac{1}{n_{i,t}} \sum_{j=1}^{n_{i,t}} SC_{i,j,t} \quad (2)$$

where $n_{i,t}$ is the number of the neighbours of node *i*.

2) TEMPORAL CORRELATION

Suppose that the behaviour of a PU is a semi-Markov ON-OFF process and all nodes in this PU's coverage range simultaneously detect the idle channels. It is clear to see that the channel availabilities of these nodes are high correlation at a timestamp, so we define the temporal correlation between node *i* and node *j* as follows:

$$TC_{i,j,t} = \frac{|c_{i,t} \cap c_{j,t}|}{n_c} \quad (3)$$

where $c_{i,t}, c_{j,t}$ are two sequences of the sensed idle channels for node *i* and node *j* respectively, and $|c_{i,t} \cap c_{j,t}|$ is the number of the common idle channels. For example, if $c_{i,t} = \{2, 3, 4\}$ and $c_{j,t} = \{2, 4\}$, then $|c_{i,t} \cap c_{j,t}| = 2$. Thus, the expectation of temporal correlation between node *i* and its one-hop neighbours is defined as follows:

$$E(TC_{i,t}) = \frac{1}{n_{i,t}} \sum_{j=1}^{n_{i,t}} TC_{i,j,t} \quad (4)$$

The larger expectation value reveals that a node and its neighbours have more common idle channels (i.e., higher temporal correlation), which makes them have high probability in a cluster.

3) CONFIDENCE LEVEL

As discussed, without shadowing problem or unsuspecting interference, the channel availabilities of the geographically close nodes have high spatial and temporal correlation so that the sensing confidences of these nodes also have high correlation. To evaluate all nodes' initial confidences before the first round of clustering, the BS broadcasts a *Training start* message. After receiving the message, each node starts to detect spectrum and sends sensing result to BS in multi-hop manner. Then, the BS makes a global decision by the majority criterion [40] and broadcasts its decision. In the *Training* stage, a confidence level is defined to measure the similarity of sensing results between node *i* and BS

$$CL_{i,0} = \frac{|c_{i,0} \cap c_{g,0}|}{|c_{g,0}|} \quad (5)$$

where $c_{i,0}, c_{g,0}$ are two sequences of the sensed idle channels for node i and BS, respectively. In the *Clustering* stage, node i updates the confidence level by exchanging the sensing results with its cluster CH_k , i.e.,

$$CL_{i,t} = \frac{|c_{i,t} \cap c_{k,t}|}{|c_{k,t}|} \quad (6)$$

where $c_{i,t}, c_{k,t}$ are two sequences of the sensed idle channels for node i and CH_k , respectively. We thus define the expectation of confidence level for node i as follows:

$$E(CL_{i,t}) = \frac{1}{T} \sum_{t=1}^T CL_{i,t} \quad (7)$$

where T is the total rounds of clustering.

4) RESIDUAL ENERGY LEVEL

One of the main challenges of spectrum-aware clustering is high energy consumption especially when exchanging sensing results and making local spectrum decision. Such issue becomes more challenging for the energy limited CRSNs, so the energy-efficient clustering is necessary to prolong network lifetime [41], [42]. Because the CHs need more energy to make local decision and allocate the sensed idle channels for member nodes, a node with higher residual energy is more likely to become a CH. To evaluate the energy of all nodes, the residual energy level for node i is defined as

$$REL_{i,t} = e_{i,t}/e_{\max} \quad (8)$$

where $e_{i,t}$ is the residual energy of node i , and e_{\max} is the maximal initial energy within all nodes.

III. SPECTRUM-AWARE CLUSTERING ALGORITHM BASED ON WEIGHTED CLUSTERING METRIC

A. CLUSTERHEAD ELECTION

The proposed evaluation criteria allow to clustering nodes involving various influence factors, which contributes to obtain the optimal spectrum-aware clustering. Now, we apply these evaluation criteria to define the following weighted clustering metric for node i

$$WCM_{i,t} = w_1(E(SC_{i,t}) \cdot E(TC_{i,t})) + w_2E(CL_{i,t}) + w_3REL_{i,t} \quad (9)$$

where $E(SC_{i,t}) \cdot E(TC_{i,t})$ represents the temporal-spatial correlation between node i and its one-hop neighbours, and w_1, w_2, w_3 are three weight coefficients indicating the influence of three influence factors on the clustering metric.

Using WCM to elect CHs and form clusters, the clustering process is described as follows. First, each node senses spectrum and exchanges the sensing information with its one-hop neighbours, followed by computing WCM and sending a WCM message to its neighbours. Then, after receiving multiple WCM messages, each node compares its WCM with all received messages. If the values of all received WCM s are less than its own, the node declares itself as a CH and broadcasts a CH Announcement message; otherwise, the node

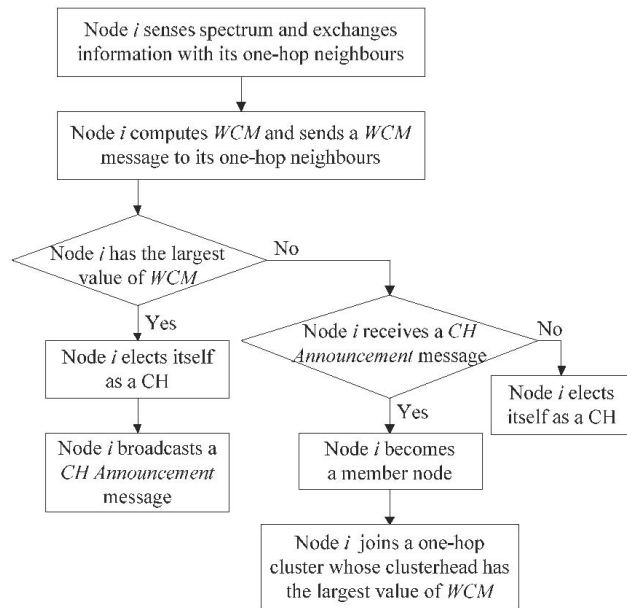


FIGURE 2. Flow chart of spectrum-aware clustering based on weighted clustering metric.

that cannot declare itself as a CH joins a one-hop cluster, whose clusterhead has the largest value of WCM . The node that does not receive any CH Announcement message elects itself as a CH. Finally, the N_s nodes are partitioned into K clusters by self-organizing mode. Figure 2 shows the flow chart of spectrum-aware clustering based on weighted clustering metric.

B. OPTIMAL NUMBER OF CLUSTERS

For cluster-based CRSN, the cluster size (i.e., the number of nodes in a cluster) has great impact on cooperative tasks such as channel allocation, channel switch and data routing. The small cluster size (i.e., the large number of clusters) inevitably incurs high routing overhead and transmission delay. On the other hand, the large cluster size (i.e., the small number of clusters) leads to less common idle channels in a cluster. Therefore, it is very important to investigate whether grouping N_s nodes into K clusters due to WCM is reasonable. In other words, whether is the number of clusters K optimal?

To solve the optimal number of clusters, we need to define the intra-cluster similarity and the inter-cluster dissimilarity. Suppose that the k -th cluster at the t -th round has a clusterhead CH_k and m_k member nodes. The intra-cluster similarity is defined by the average temporal-spatial correlation between CH_k and its member nodes

$$ICR_{k,t} = \begin{cases} \frac{1}{m_k} \sum_{i=1}^{m_k} SC_{i,k,t} \cdot TC_{i,k,t}, & m_k > 0 \\ 0, & m_k = 0 \end{cases} \quad (10)$$

where $SC_{i,k,t} \cdot TC_{i,k,t}$ represents the temporal-spatial correlation between the i -th member nodes and CH_k . For convenience, we use the temporal-spatial correlation between

two CHs to denote the correlation between two clusters. The inter-cluster dissimilarity is thus defined by the average temporal-spatial incorrelation between the k -th cluster and the other $K - 1$ clusters

$$UNR_{k,t} = \frac{1}{K-1} \sum_{l=1, l \neq k}^K (1 - SC_{l,k,t} \cdot TC_{l,k,t}) \quad (11)$$

where $1 - SC_{l,k,t} \cdot TC_{l,k,t}$ represents the temporal-spatial incorrelation between the l -th cluster and the k -th cluster. Combining (10) and (11), the following optimization model maximizing the intra-cluster similarity and the inter-cluster dissimilarity is proposed to solve the optimal number of clusters K^*

$$\max_K L(K) = \frac{1}{K} \sum_{k=1}^K IRA_{k,t} + \frac{1}{K} \sum_{k=1}^K UNR_{k,t} \quad (12)$$

Note that our proposed algorithm has grouped N_s nodes into K clusters so that we can compute the corresponding objective function $L(K)$. As we all know, some of clusters with high temporal-spatial correlation can be merged while some of clusters with low temporal-spatial correlation usually remain separation. Therefore, we merge the pair of clusters with the highest temporal-spatial correlation, where one of two CHs with larger value of WCM is elected as a new CH. Next, the objective function $L(K - 1)$ is calculated for the new $K - 1$ clusters. If $L(K - 1)$ is smaller than $L(K)$, the existing K clusters is optimal and the network topology remains unchanged. Otherwise, we sequentially merge clusters until the corresponding value of objective function no longer becomes larger. By this way, the N_s nodes are grouped into clusters with the optimal number of clusters K^* .

From the above discussion, our proposed algorithm solving the optimal number of clusters is summarized as follows.

After clustering, the BS constructs a shortest path tree routing for all CHs according to their reported locations, and then the CRSN steps into the *Data Transmission* stage. In intra-cluster range, each CH broadcasts a *Transmission start* packet containing the channel assignment message and the transmission time. Each member node transmits data to its CH after it receives the *Data Transmission start* packet. In inter-cluster range, all CHs transmit the collected data to BS by the established routing. In the interval of data transmission, all CHs periodically sense spectrum instead of all member nodes. Once the channel availability changes in the monitoring area, the CRSN will trigger the local or global re-clustering to establish a new optimal clustering in terms of the CHs' sensing results. As such, the high-efficiency spectrum monitoring based on the optimal clustering is a key to implement high-performance dynamic spectrum sharing and high-speed data transmission.

IV. EXPERIMENTAL RESULTS AND ANALYSES

In this section, the feasibility and effectiveness of our proposed algorithm are studied and analyzed by the following extended simulation experiments in MATLAB platform.

Algorithm 1 The Optimal Number of Clusters for Spectrum-Aware Clustering Algorithm Based on Weighted Clustering Metric

Initialization: At the t -th round, group nodes into K clusters by comparing their WCM s

- 1: while $K > 2$ do
- 2: for each cluster $k \in \{1, \dots, K\}$
- 3: calculate $ICR_{k,t}$ and $UNR_{k,t}$;
- 4: end
- 5: calculate $L(K)$;
- 6: find one pair clusters

$$(l_i, l_j) = \max SC_{l_i, l_j, t} \cdot TC_{l_i, l_j, t};$$
- 7: merge two clusters and elect a new clusterhead

$$Cluster C_i \leftarrow Cluster C_i \cup Cluster C_j;$$
- 8: for each cluster $k' \in \{1, \dots, K - 1\}$
- 9: calculate $ICR_{k',t}$ and $UNR_{k',t}$;
- 10: end
- 11: calculate $L(K - 1)$;
- 12: if $L(K - 1) \leq L(K)$
- 13: $K^* = K$;
- 14: else if $L(K - 1) > L(K)$
- 15: $K = K - 1$;
- 16: end
- 17: end

Output: the optimal number of clusters K^* .

A. PERFORMANCE ANALYSIS OF SPECTRUM-AWARE CLUSTERING ALGORITHM

We consider a regular cognitive radio sensor network where the PUs and the CSNs share radio spectrum by CSNs sensing and accessing the PUs' idle channels. The experimental scenario is set as follows: 3 PUs and 50 sensor nodes are randomly distributed in a square area with size $200m^2$, and their communication ranges are 100m and 50m, respectively. The spectrum resource in this area is divided into 8 channels licensed to PU_1, PU_2, PU_3 . In Figure 3, three PUs are occupying the channels $\{2,3\}, \{2,5\}$ and $\{1,6\}$, respectively. In other word, the idle channels that the nodes can detect are $\{1,4,5,6,7,8\}, \{1,3,4,6,7,8\}$ and $\{2,3,4,5,7,8\}$, respectively. Accordingly, we set the sensed idle channels of a node due to its geographical location. For example, if a node is in the coverage ranges of PU_1 and PU_2 , its sensed idle channels are randomly chosen from $\{1,4,6,7,8\}$. The remaining energy level of each node is randomly chosen from $(0,1)$. In terms of the sensing results of all nodes, Figure 3(a) displays that the typical spectrum-aware clustering algorithm groups nodes into 13 clusters, where 5 nodes cannot join any cluster and have to declare themselves as clusterheads $CH_9 \sim CH_{13}$. These nodes, however, may group together due to the temporal-spatial correlation. Setting $w_1 = w_2 = w_3 = 1/3$, our proposed algorithm divides 50 nodes into 10 clusters in Figure 3(b), where only three nodes declare themselves as clusterheads CH_8, CH_9, CH_{10} . It is clear to see that grouping nodes based on various influence factors

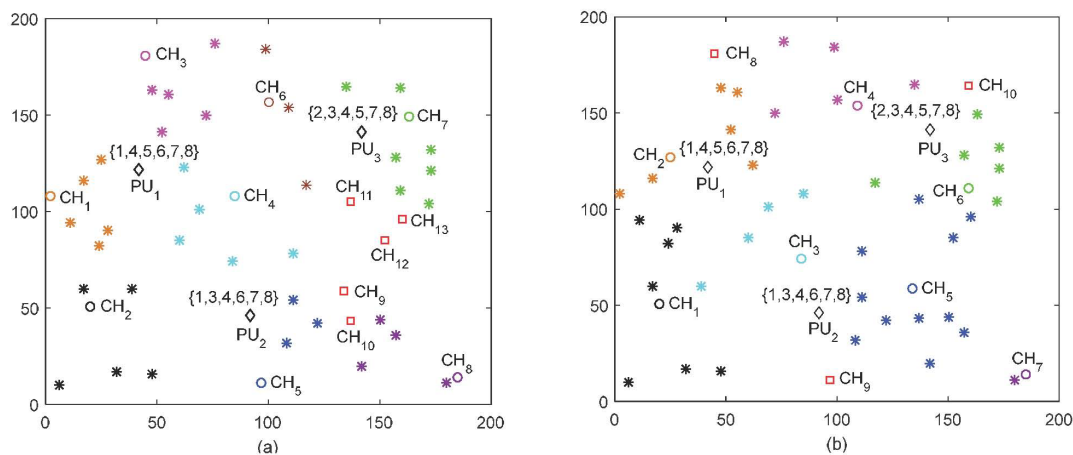


FIGURE 3. Comparison of clustering results for different spectrum-aware clustering algorithms. (a) Clustering using typical spectrum-aware clustering algorithm (b) Clustering using our proposed algorithm.

can provide more reasonable spectrum-aware clustering compared to only focus on the sensed idle channels of all nodes.

B. VERIFICATION OF THE OPTIMAL CLUSTERING

We next try to verify our spectrum-aware clustering is optimal. Applying Algorithm 1 to the clustering result in Figure 3(b), Figure 4 shows that the objective function $L(K)$ is monotone decreasing with respect to the number of clusters K , where the maximum function value is obtained as $K = 10$. More specifically, merging a pair clusters C_5, C_7 with the maximum temporal-spatial correlation, the objective function value of the new clustering $L(9)$ reduces 1.97% compared to $L(10)$. Nevertheless, the new clusterhead CH_5 elected from $\{CH_5, CH_7\}$ cannot communicate with the member nodes in the cluster C_7 , so the merge operation cannot obtain better clustering than the initial clustering with 10 clusters. The similar result happens after merging the new generated cluster C_5 and the previous cluster C_6 . Repeating the above merge operation, the value of objective function decreases until all nodes finally form two clusters. The monotone decreasing property of the objective function

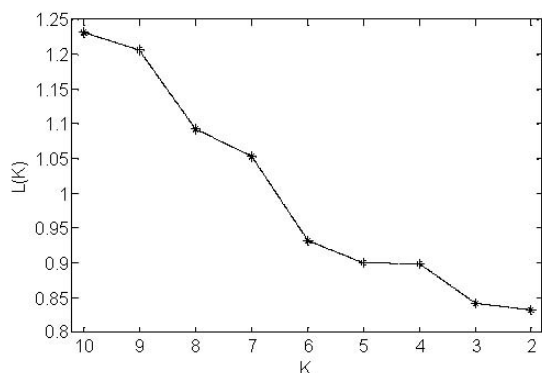


FIGURE 4. Objective function $L(K)$ with respect to the number of clusters K .

implies that our proposed algorithm can provide the optimal spectrum-aware clustering when considering the intra-cluster and inter-cluster correlation characteristics.

C. EVALUATION FOR NETWORK PERFORMANCE

After forming the optimal clustering, the member nodes receive the *Data Transmission start* packet from their CHs and start to transmit data. To test the cluster-based network performance, the following simulation experiments set the system parameters based on IEEE 802.15.4 as follows. We perform a round of the PU's ON-OFF process, where each process maintains 60s. Due to the new sensing structure in Figure 1, the sensing time and sensing period are 10ms and 40ms, respectively, and the re-clustering costs 40ms. The rate of each node's data generation is 2kb/s, and the messages for sensing and clustering are 150bits and 300bits, respectively. The energy consumption for handling these messages is 50nJ/bit. During the available time of data transmission, each member node transmits sensing data to its CH where the energy consumption of data transmission follows the free space transport model [28]. Each CH then fuses its own data and the received data with the data fusion rate 20%. Finally, the BS located on (150, 0) fuses the received data from all CHs to accomplish data collection and information management. In Figure 5(a), our proposed algorithm only needs CHs to periodically sense spectrum after clustering that helps their member nodes achieve more data transmission time to improve the network throughput. On the contrary, the periodical spectrum sensing algorithm in [18] needs all nodes periodically sense spectrum, which makes the amount of data transmission reduces from 8.28% to 21.63% with the increasing number of nodes compared to our proposed algorithm. On the other hand, Figure 5(b) displays that the energy consumption of spectrum sensing and clustering for our proposed algorithm is less than that of the periodical spectrum sensing algorithm, where the maximum reduction

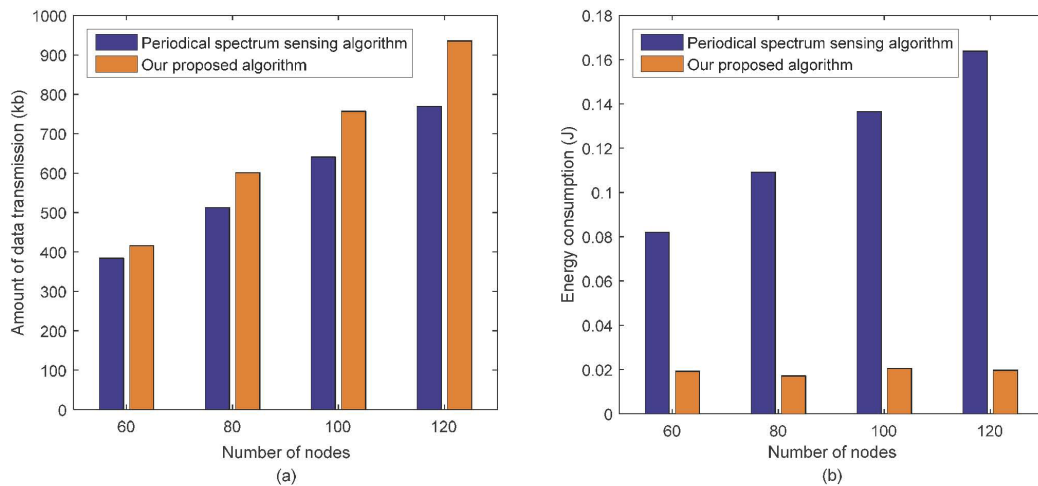


FIGURE 5. Comparison of the network performance for the periodical spectrum sensing algorithm and our proposed algorithm after clustering. (a) Amount of data transmission for different number of nodes (b) Energy consumption of spectrum sensing and clustering for different number of nodes.

is about 88% as $N_s = 120$. The above experimental results reveal that applying the new sensing structure makes our proposed algorithm obtain better network performance.

D. INFLUENCE OF THREE WEIGHT COEFFICIENTS

From the implement processes of our spectrum-aware clustering algorithm, we observe that the *WCM* plays an important role in clusterhead election and cluster formation. Furthermore, three weight coefficients w_1, w_2, w_3 are the influence of temporal-spatial correlation, sensing confidence and residual energy on *WCM*, respectively. In the front experiments, we suppose that three weight coefficients have the same value, i.e., three evaluation factors have the equal influence on *WCM*. In different practical scenarios, however, the importance of three evaluation factors to spectrum-aware clustering is different. For instance, with the purpose of real-time data transmission, the clustering may mainly think about sensing confidence because highly reliable nodes cooperatively detecting idle channels is the guarantee of high-quality communication. On the other hand, in a video monitoring system, the CHs must have residual energy as much as possible in order to fuse a large amount of the video signals observed by their member nodes. At this point, the clustering needs to emphasize on the effect of residual energy on *WCM*.

The following experiments test the performance of our proposed algorithm with respect to three weight coefficients according to the different application scenarios. To achieve the most compact hierarchical structure for a CRSN, we set $w_1 = 3/4, w_2 = 1/8, w_3 = 1/8$ to emphasize on the influence of temporal-spatial correlation. Figure 6(a) shows that 50 nodes are grouped into 9 clusters, where only one node elects itself as CH. Compared to the equal weight coefficients in Figure 3(b), *WCM* giving priority to temporal-spatial

correlation can effectively clustering nodes thereby avoiding too much isolated CHs. When weakening the influence of temporal-spatial correlation and reinforcing the influence of sensing confidence, two nodes elect themselves as CHs in Figure 6(b). Figure 6(c) gives the most number of clusters if emphasizing on the influence of residual energy. Therefore, the results of spectrum-aware clustering mainly depends on the linear combination mode of three evaluation factors in *WCM*.

To evaluate the network performance for the different linear combination modes of three evaluation factors, we define the average amount of data transmission and the average energy consumption by performing Monte Carlo experiment. Following the above experiments, Figure 7(a) shows that the average number of clusters for four linear combination modes greatly increases with the increasing of the number of nodes, which leads to the notably increasing of the average energy consumption in Figure 7(c). The average energy consumption with respect to $w_1 = 1/8, w_2 = 1/8, w_3 = 3/4$ is higher than that of the other linear combination modes, and the average amount of data transmission is the lowest in Figure 7(b). This is because that spectrum-aware clustering mainly focusing on the effect of residual energy cannot take advantage of the high correlation of sensing results to form reasonable clustering. The other linear combination modes, however, make use of temporal-spatial correlation to obtain better network performance. Especially when setting $w_1 = 3/4, w_2 = 1/8, w_3 = 1/8$, we have the best network performance including the smallest number of clusters, the highest average amount of data transmission and the lowest average energy consumption. It is easy to see that the temporal-spatial correlation plays a more important role on the spectrum-aware clustering, which is consistent with the previous theoretical analyses. Our clustering metric well utilizes this property to achieve the satisfactory spectrum-aware clustering-based

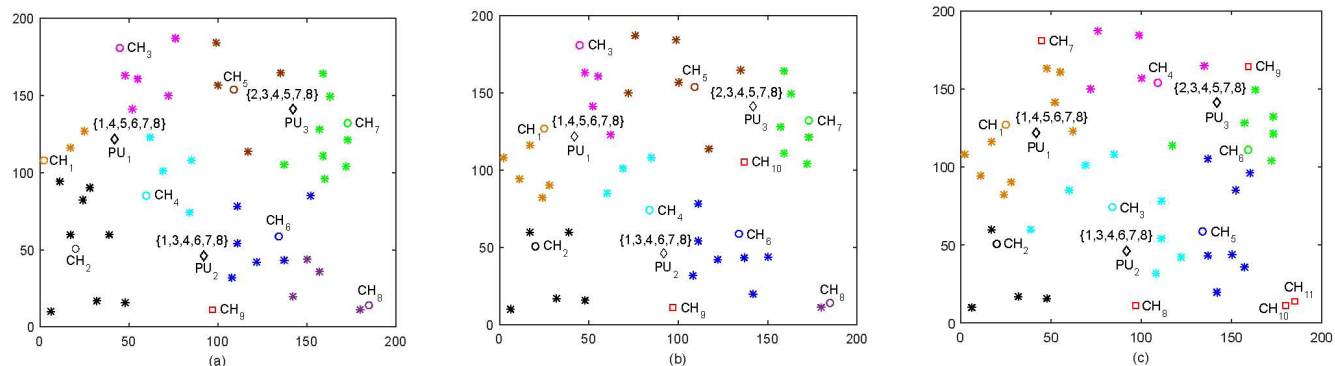


FIGURE 6. Comparison of spectrum-aware clustering for different linear combination modes of three weight coefficients. (a) $w_1 = 3/4, w_2 = 1/8, w_3 = 1/8$ (b) $w_1 = 1/8, w_2 = 3/4, w_3 = 1/8$ (c) $w_1 = 1/8, w_2 = 1/8, w_3 = 3/4$.

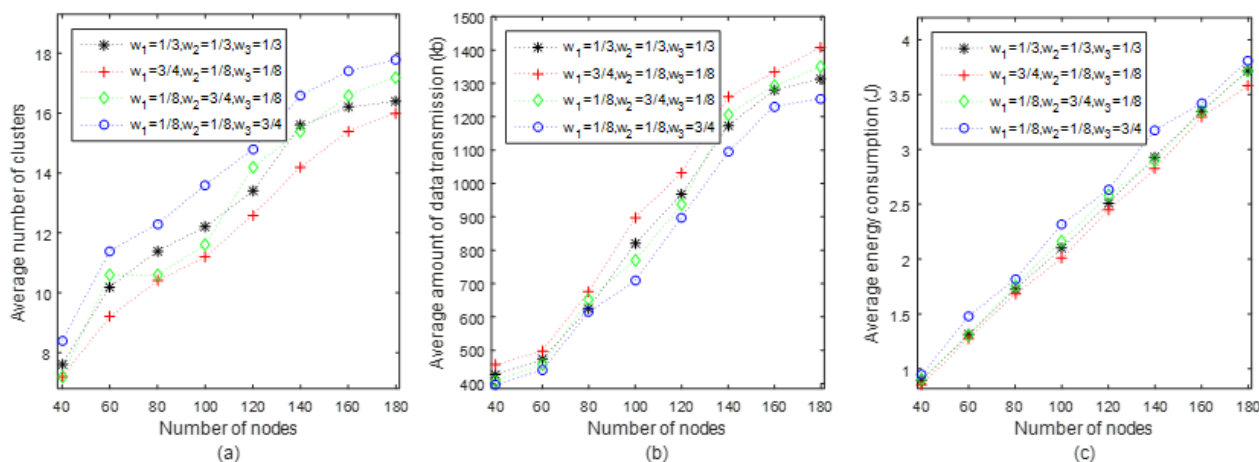


FIGURE 7. Comparison of the network performance for the different values of three weight coefficients. (a) Number of clusters for the different number of nodes (b) Amount of data transmission for the different number of nodes (c) Average energy consumption of data transmission for the different number of nodes.

network topology so as to improve the quality of service of CRSNs.

E. PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

The next experiment compares the performance of our proposed algorithm with two existing spectrum-aware clustering algorithms in terms of the number of clusters. Figure 8 illustrates the relation between the number of nodes and the number of clusters for three algorithms. It can be seen that the number of clusters for three algorithms increases with the increasing of the number of nodes, where the proposed algorithm using $w_1 = w_2 = w_3 = 1/3$ constructs the less number of clusters. For example, when the number of nodes up turns to 500, the proposed algorithm groups nodes into 22 clusters while RARE and cluster-based protocol [36] create 24 clusters and 26 clusters, respectively. Obviously, the high node density enhances the temporal-spatial correlation that makes a cluster include the more number of neighboring nodes

in our proposed algorithm. In terms of the experimental conditions in Figure 7, Figure 8(b) displays the average energy consumption of data transmission for three clustering algorithms after forming clusters listed in Figure 8(a). The average energy consumption of RARE and our proposed algorithm are lower than that of cluster-based protocol, and the corresponding reduction proportion is from 3.45% to 51.74%. This is because the number of clusters of RARE and our proposed algorithm are less than that of cluster-based protocol. The energy efficiency of our proposed algorithm, which forms the optimal clustering, outperforms RARE and cluster-based protocol. As $N_s = 500$, the average energy consumption of our proposed algorithm reduces 3.12% and 9.81% compared to RARE and cluster-based protocol. From the experiments in Figure 7 and Figure 8, we can clearly see that the temporal-spatial similarity of spectrum sensing is the essential characteristic of CRSNs. Taking full advantage of this characteristic is not only benefit to quickly form clusters, but also to design effective routing protocol and data transmission protocol.

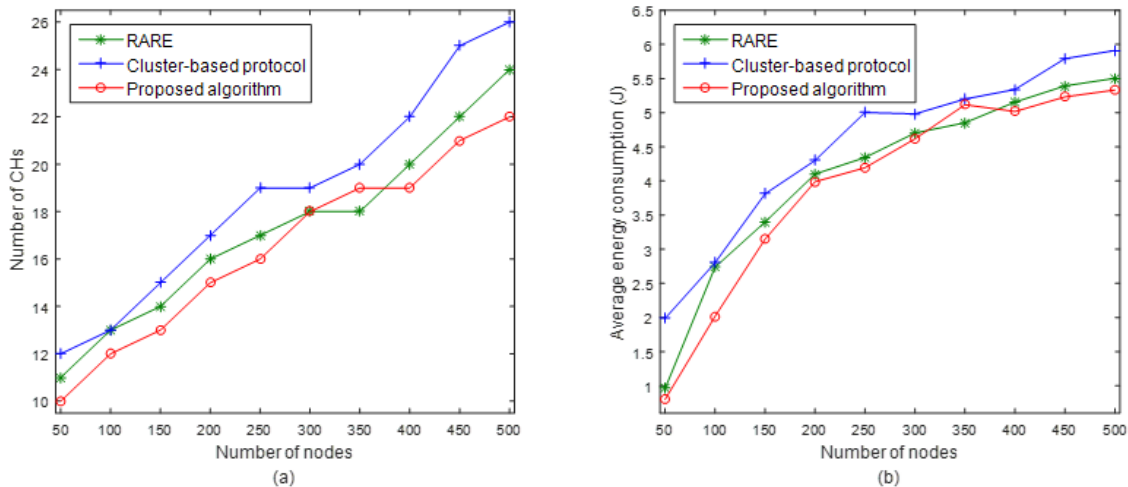


FIGURE 8. Performance comparison of the proposed algorithm with RARE and cluster-based protocol. (a) Number of clusters for the different number of nodes (b) Average energy consumption of data transmission for the different number of nodes.

V. CONCLUSION

The spectrum-aware clustering algorithm can better group nodes than some typical clustering algorithms by well exploiting the cognitive characteristic of CRSNs. Different from the traditional spectrum-aware clustering algorithms, a novel spectrum-aware clustering algorithm based on weighted clustering metric is proposed to take account into various clustering criteria including temporal-spatial correlation, sensing confidence and residual energy. Comparing the WCM of all nodes, the optimal spectrum-aware clustering is formed to efficiently enhance the network connectivity of a CRSN. After clustering, our proposed algorithm involving CHs sensing spectrum instead of all member nodes provides an energy-efficient spectrum sensing method. The experiments show that the data transmission and energy consumption are greatly improved under the optimal spectrum-aware clustering. In the future work, we will test the proposed algorithm in some application scenarios, e.g., smart community, internet of vehicles and ecological monitoring.

REFERENCES

- [1] S. Bukhari, S. Siraj, and M. H. Rehmani, "PRACB: A novel channel bonding algorithm for cognitive radio sensor networks," *IEEE Access*, vol. 4, pp. 6950–6963, 2016.
- [2] E. U. Ogbodo, D. Dorrell, and A. M. Abu-Mahfouz, "Cognitive radio based sensor network in smart grid: Architectures, applications and communication technologies," *IEEE Access*, vol. 5, pp. 19084–19098, 2017.
- [3] R. Samir, M. S. E. Mahallawy, S. M. Gasser, and N. Zaher, "Exploring the effect of various cluster structures on energy consumption and end-to-end delay in cognitive radio wireless sensor networks," *IEEE Access*, vol. 6, pp. 38062–38070, 2018.
- [4] A. Ahmad, S. Ahmad, M. H. Rehmani, and N. U. Hassan, "A survey on radio resource allocation in cognitive radio sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 888–917, 2nd Quart., 2015.
- [5] M. Usman, D. Har, and I. Koo, "Energy-efficient infrastructure sensor network for ad hoc cognitive radio network," *IEEE Sensors J.*, vol. 16, no. 8, pp. 2775–2787, Apr. 2016.
- [6] H. Wu, F. Yao, Y. Chen, Y. Liu, and T. Liang, "Cluster-based energy efficient collaborative spectrum sensing for cognitive sensor network," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2722–2725, Oct. 2017.
- [7] W. Zhang, Y. Yang, and C. K. Yeo, "Cluster-based cooperative spectrum sensing assignment strategy for heterogeneous cognitive radio network," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2637–2647, Jun. 2015.
- [8] V. D. Nguyen, O. T. T. Kim, D. N. M. Dang, S. S. Kim, and C. S. Hong, "Application of the lowest-ID algorithm in cluster-based TDMA system for VANETs," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, 2015, pp. 25–30.
- [9] S. K. Gupta, N. Jain, and P. Sinha, "Node degree based clustering for WSN," *Int. J. Comput. Appl.*, vol. 40, no. 16, pp. 49–55, 2012.
- [10] Z. Fareeha and M. Zaigham, "Comparative analysis of clustering algorithms comprising GESC, UDCA, and k-Mean methods for wireless sensor networks," *URSI Radio Sci. Bull.*, vol. 339, pp. 12–18, Dec. 2011.
- [11] Asaduzzaman and H. Y. Kong, "Energy efficient cooperative LEACH protocol for wireless sensor networks," *J. Commun. Netw.*, vol. 12, no. 4, pp. 358–365, Aug. 2010.
- [12] N. R. Roy and P. Chandra, "A note on optimum cluster estimation in LEACH protocol," *IEEE Access*, vol. 6, pp. 65690–65696, 2018.
- [13] R. M. Eletreby, H. M. Elsayed, and M. M. Khairy, "CogLEACH: A spectrum aware clustering protocol for cognitive radio sensor networks," in *Proc. CROWNCOM*, Jun. 2014, pp. 179–184.
- [14] H. Z. Han, Z. H. Sun, B. Shen, and T. Q. Zhang, "LEAUCH: Low-energy adaptive uneven clustering hierarchy for cognitive radio sensor network," *EURASIP J. Wireless Commun. Netw.*, vol. 122, pp. 1–122, Dec. 2015.
- [15] N. Mansoor, A. K. M. M. Islam, M. Zareei, and C. Vargas-Rosales, "RARE: A spectrum aware cross-layer MAC protocol for cognitive radio ad-hoc networks," *IEEE Access*, vol. 6, pp. 22210–22227, 2018.
- [16] B. Shahrabi and N. Rahnavard, "A clustering-based coordinated spectrum sensing in wideband large-scale cognitive radio networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 1101–1106.
- [17] I. Pefkianakis, S. H. Y. Wong, and S. W. Lu, "Spectrum-aware Location-based Routing in Cognitive UWB network," in *Proc. 3rd IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2008, pp. 14–17.
- [18] A. Ali, I. Yaqoob, E. Ahmed, M. Imran, K. S. Kwak, A. Ahmad, S. A. Hussain, and Z. Ali, "Channel clustering and QoS level identification scheme for multi-channel cognitive radio networks," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 164–171, Apr. 2018.
- [19] Y. Saleem, K. L. A. Yau, H. Mohamad, N. Ramli, and M. H. Rehmani, "SMART: A spectrum-aware cluster-based routing scheme for distributed cognitive radio networks," *Comput. Netw.*, vol. 91, pp. 196–224, Nov. 2015.
- [20] N. Dutta, H. K. D. Sarma, and Z. Polkowski, "Cluster based routing in cognitive radio adhoc networks: Reconnoitering SINR and ETT impact on clustering," *Comput. Commun.*, vol. 115, pp. 10–20, Jan. 2018.
- [21] D. Li, E. Fang, and J. Gross, "Robust clustering for ad hoc cognitive radio network," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 5, 2018, Art. no. e3285.
- [22] N. Mansoor, A. M. Islam, M. Zareei, S. Baharun, and S. Komaki, "A novel node joining algorithm for spectrum aware cluster-based cognitive radio ad-hoc networks," in *Proc. Control Conf.*, May 2015, pp. 1–6.
- [23] X. Y. Li, Q. K. Xiao, and Y. Zhang, "Clustering algorithm for cognitive radio," *Boletín Técnico*, vol. 55, pp. 491–497, 2017.

[24] M. Bradonjić and L. Lazos, "Graph-based criteria for spectrum-aware clustering in cognitive radio networks," *Ad Hoc Netw.*, vol. 10, no. 1, pp. 75–94, 2012.

[25] M. Ozger, F. Alagoz, and O. B. Akan, "Clustering in multi-channel cognitive radio ad hoc and sensor networks," *IEEE Commun. Mag.*, vol. 56, no. 14, pp. 156–162, Apr. 2018.

[26] H. Zhang, Z. Zhang, H. Dai, R. Yin, and X. Chen, "Distributed spectrum-aware clustering in cognitive radio sensor networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–6.

[27] Z. Qu, Y. Xu, and S. Yin, "A novel clustering-based spectrum sensing in cognitive radio wireless sensor networks," in *Proc. Cloud Comput. Intell. Syst. (CCIS)*, Nov. 2015, pp. 695–699.

[28] F. A. Fsana, N. Jahan, F. A. Sunny, M. S. Kaiser, and S. A. Mamun, "Trust and energy aware cluster modeling and spectrum handoff for cognitive radio ad-hoc network," in *Proc. Elect. Eng. Inf. Commun. Technol. (ICEE-ICT)*, May 2015, pp. 1–6.

[29] L. De Nardis, D. Domenicali, and M.-G. Di Benedetto, "Clustered hybrid energy-aware cooperative spectrum sensing (CHESS)," in *Proc. Cogn. Radio Oriented Wireless Netw. Commun.*, Jun. 2009, pp. 1–6.

[30] K. Souvik, S. Karthikeyan, and A. Karthikeyan, "QBEECH: Multi-hop clustering of cognitive based sensor nodes in the administration of queen nodes," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2018, pp. 376–385.

[31] I. Mustapha, B. M. Ali, M. F. A. Rasid, A. Sali, and H. Mohamad, "An energy-efficient spectrum-aware reinforcement learning-based clustering algorithm for cognitive radio sensor networks," *Sensors*, vol. 15, pp. 19783–19818, Aug. 2015.

[32] A. Latiwesh and D. Qiu, "Energy efficient spectrum aware clustering for cognitive sensor networks: CogLEACH-C," in *Proc. 10th Int. Conf. Commun. Netw. China*, Aug. 2015, pp. 515–520.

[33] H. Z. Zhang, Z. Y. Zhang, and C. Yuen, "Energy-efficient spectrum-aware clustering for cognitive radio sensor networks," *Chin. Sci. Bull.*, vol. 57, nos. 28–29, pp. 3731–3739, 2012.

[34] Y. R. Narayan, M. Rajiv, and S. Divya, "Energy aware cluster based routing protocol over distributed cognitive radio sensor network," *Comput. Commun.*, vol. 129, pp. 54–66, Sep. 2018.

[35] S. Kumar and A. K. Singh, "A localized algorithm for clustering in cognitive radio networks," *J. King Saud Univ.-Comput. Inf. Sci.*, to be published.

[36] X. Li, F. Hu, H. Zhang, and X. Zhang, "A cluster-based MAC protocol for cognitive radio ad hoc networks," *Wireless Pers. Commun.*, vol. 69, no. 2, pp. 937–955, 2013.

[37] M. Zheng, S. Chen, W. Liang, and M. Song, "NSAC: A novel clustering protocol in cognitive radio sensor networks for Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5864–5865, Feb. 2019.

[38] S. Atapattu, C. Tellambura, H. Jiang, and N. Rajatheva, "Unified analysis of low-SNR energy detection and threshold selection," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5006–5019, Nov. 2015.

[39] A. Ali and W. Hamouda, "Advances on spectrum sensing for cognitive radio networks: Theory and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1277–1304, 2nd Quart., 2016.

[40] X. Huang, H. B. Cheng, and G. Yang, "Research of connectivity for wireless sensor networks," *J. Commun.*, vol. 30, no. 2, pp. 129–135, 2009.

[41] R. Zhu, X. Zhang, X. Liu, W. Shu, T. Mao, and B. Jalaian, "ERDT: Energy-efficient reliable decision transmission for intelligent cooperative spectrum sensing in industrial IoT," *IEEE Access*, vol. 3, pp. 2366–2378, 2015.

[42] C. Wang, J. Wu, Y. Yu, T. Song, and J. Hu, "Energy-efficient cooperative spectrum sensing with reporting errors in hybrid spectrum sharing CRNs," *IEEE Access*, vol. 6, pp. 48391–48402, 2018.



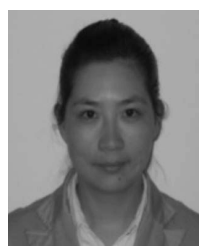
XINJIE GUAN received the B.S. and M.S. degrees from Southeast University, in 2006, and the Ph.D. degree from the University of Missouri-Kansas City, in 2015, all in computer science. Since 2016, she has been an Assistant Professor with the School of Computer Science and Technology, Nanjing Tech University, China. Her research interests include network optimization, cloud computing, and mobile edge computing.



XILI WAN received the B.S. and M.S. degrees from Nanjing Tech University, and the Ph.D. degree from the University of Missouri-Kansas City, USA, in 2014, all in computer science. He was Researcher with Huawei USA Research Center. Since 2016, he has been an Assistant Professor with the School of Computer Science and Technology, Nanjing Tech University, China. His research interests include optimization in wireless sensor networks, design and analysis of approximation algorithms, and cloud computing.



HANG SHEN received the B.S., M.S., and Ph.D. degrees in computer science from Nanjing Tech University, where he has been an Assistant Professor with the School of Computer Science and Technology, since 2015. His research interests include wireless networks, multimedia, data privacy, and network coding.



TIANJING WANG received the B.S. degree in mathematics from Nanjing Normal University, in 2000, the M.S. degree in computing mathematics from Nanjing University, in 2005, and the Ph.D. degree in signal and information processing from the Nanjing University of Posts and Telecommunications, in 2009. From 2010 to 2013, she was with the Postdoctoral Research Center, Nanjing University of Posts and Telecommunications. Since 2009, she has been an Assistant Professor with

the School of Computer Science and Technology, Nanjing Tech University, China. Her main research interests include signal processing, wireless networks, and optimal method.



XIAOMEI ZHU received the B.S. and M.S. degrees from the China University of Mining and Technology, and the Ph.D. degree from the Nanjing University of Posts and Telecommunications, all in signal processing. Since 2009, she has been an Assistant Professor with the School of Computer Science and Technology, Nanjing Tech University, China. Her research interests include signal detection and estimation, spectrum sensing, and deep learning for modern communications systems.

...