Received November 21, 2019, accepted December 6, 2019, date of publication December 13, 2019,
date of current version December 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959313

Q-Learning Based Multi-Objective Clustering Algorithm for Cognitive Radio Ad Hoc Networks

MD ARMAN HOSSEN AND SANG-JO YOO, (Member, IEEE)
Department of Information and Communication Engineering, Inha University, Incheon 402-751, South Korea

Corresponding author: Sang-Jo Yoo (sjyoo@inha.ac.kr)

This work was supported by the Inha University Research Grant.

ABSTRACT Cognitive radio (CR) is an adaptive radio technology that can automatically detect available channels in a wireless spectrum and change transmission parameters to improve radio operating behavior. Due to the dynamic nature of spectrum availability and wireless channel condition, it is very hard to maintain reliable network connectivity. Cluster-based CR ad-hoc networks (CRAHN) arrange CR nodes into groups to effectively maintain reliable autonomous networks. Clustering in CRAHN supports cooperative tasks such as spectrum sensing and channel managements and achieves network scalability and stability. In this paper, we proposed a Q-learning based cluster formation approach in CRAHN, in which Q-value is used to evaluate each node's channel quality. To form a distributed cluster network, channel quality, residual energy and neighbor node/network conditions are considered. By exchanging each node's status information in terms of channels and neighbors, each node knows neighboring topology and which node is the best candidate for cluster head (CH). Distributed CH selection, the optimum common active data channel decision, and gateway node selection procedures are presented in this paper. The proposed mechanism can extend the network lifetime, enhance the reachability not only between member nodes but also with other cluster networks, it can also provide stable and reliable service using the selected data channel and avoid possible interference between neighboring ad-hoc clusters.

INDEX TERMS Reinforcement learning, clustering, cognitive radio, Q-learning, ad-hoc network.

I. INTRODUCTION

The term cognition has come from a Latin phrase cognoscere, means getting to know, or knowledge. Cognitive radio (CR) has been the propitious field for future generation wireless communication. In the cognitive radio system, the secondary users (SUs) can explore and exploit any licensed spectrum which is owned by primary users (PUs) without causing any harmful interference to the PUs. In the era of excessive demand for wireless communication, with the scarcity of licensed spectrum, CR is the promising choice to expand the communication systems and provides the proper utilization of spectrum. With the limitation of licensed spectrum, the Federal Communications Commission (FCC) has endorsed the licensed bands available for the unlicensed devices when it's free [1]. IEEE has a standard for CR networks, IEEE 802.22, designed for a centralized network where CR devices use the white space of TV frequency bands. CR was first introduced by Mitola III [2]. In CR networks, SUs sense the spectrum with their dynamic spectrum access (DSA) functionality and determine the idle primary channels, decide the optimum operating channel, and vacate the channel when PUs arrive.

Considering the spectrum sharing schemes, cognitive radio networks (CRNs) can be classified into two network types, (i) a centralized network which consists of a base station (BS) and nodes where nodes send their information to the BS and BS decides most of the functions, (ii) a distributed network which consists of a number of nodes where each node communicates with others and form network without any presence of fixed infrastructure. A network can also be divided into a single channel and multi-channel aspects where nodes can utilize single or multiple primary channels, respectively. In our proposed approach to capture the dynamicity of CR nodes, we have considered the distributed CR ad hoc network (CRAHN).

Clustering is the process of making groups of nodes in geographic proximity where nodes in each group have many common features. The main advantage of clustering in CRN is providing network scalability, maintaining spectrum stability, and achieving cooperative tasks such as channel sensing

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

and access. Scalability in cluster-based CRAHN can be achieved by reducing communication range limited by cluster domain so that energy consumption and routing overhead are also reduced. Since CRN is dynamic in nature in terms of channel availability and node connectivity, if the network is maintained as a small group, then the stability is achieved through mitigating the dynamic channel assignment issue within a group.

Clustering can make further benefits in topology management via assigning control channel and enhance sensing performance. Generally, in a cluster-based CRAHN, there are cluster head (CH), gateway nodes (GNs) and member nodes (MNs). CH is the central processor which maintains the communication within a cluster (intra-cluster communication) as well as between clusters (inter-cluster communication). GNs provide communication links between two neighboring clusters. In this paper, we assume that there exists a common control channel (CCC) to exchange control messages between neighboring nodes. To form a cluster at the initialization procedure, nodes will broadcast and exchange messages using the predetermined CCC for CRN. In conventional clustering algorithms in CRN, they usually defined different objective functions such as increasing spectrum sensing accuracy, reducing energy consumption, maximizing the number of member nodes that utilize the same available channel, mitigating interference between neighbor networks, or fast reconfiguring the cluster when PU appears.

In cluster formation, the cluster head selection is the most important process because network topology, network life time, the cluster common data channel quality, connectivity with member nodes and neighbor clusters will be affected by CH selection. In this paper, we propose a Q-learning based cluster formation method for distributed CRAHN. As a reinforcement learning, Q-learning is used to evaluate and estimate channel quality at each node in terms of opportunistic channel access possibility and sustainable operation time. The Q-values for the available channels at each node are exchanged between neighbor nodes and then the proposed CH selection and common active data channel (CADC) decision processes are applied. We also propose a GN selection method that can provide effective inter-cluster connectivity. In the proposed cluster formation for CRAHN, we have defined multi-objective function that considers channel quality, network life time, even energy consumption, number of member nodes, network connectivity and coexistence with neighbor networks.

This paper is organized as follow: In Section II, we discuss related work. The proposed system architecture is presented in Section III. The Q-learning based cluster formation procedures are delineated in Section IV. The simulation results are presented in Section V and in Section VI, we conclude this paper.

II. RELATED WORK

In recent time, several clustering approaches in CR have taken place to increase network operation efficiency. Chen *et al.* [3] introduced a cluster-based framework to form a wireless mesh network in the context of open spectrum sharing for CRN, in which they proposed a decentralized cluster-based architecture to form a large-scale network. It only focused on the cluster formation but did not deeply considered the selection of active common data channel and effects of dynamic changes in clustering.

In [4], a distributed cluster-based routing algorithm is proposed for CRN for maximizing the network throughput and minimizing the end-to-end delay. Initially, nodes organize themselves into several clusters by the clustering algorithm based on location, communication efficiency, network connectivity and spectrum availability. Following completion of cluster formation, routing is done according to the spectrum usage and interference metrics. Santosh Kumar et al. proposed a weight-based clustering algorithm for CRN [5], in which each node computes its weight in terms of the number of available channels, speed of a node and PU interference level and shares with its one hop neighbors. Then a node with maximal weight becomes the cluster head. In [6], an improved k-means clustering algorithm is applied in cognitive radio ad hoc networks. K centroids are selected randomly and each node measures the distance from the centroids and updates the information with neighbors where the process continued. The closest node of a centroid is selected as CH. The problem with k-means algorithm is the number of centroids or clusters should be predefined which sometimes make the clustering process vulnerable. For example, if the number of nodes in a system is very few but centroids are defined quite close to the number of sensors, then the cluster size will be very small. The opposite case occurs when there exist lots of nodes but centroids are few.

In [7], another popular method called CogLEACH is proposed for cluster formation for CR sensor networks, which is a spectrum-aware extension of the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol. CogLEACH uses the number of vacant channels as a weight in the probability of each node to become a CH. Here inter-cluster communication is maintained through Direct Sequence Spread Spectrum (DSSS) spreading codes which are secure but comparatively time worthy and complex. In [8], a robust clustering algorithm for CRAHN is proposed, in which the spatial variations of spectrum availability are considered for clustering. A parameter named Cluster Head Determination Factor (CHDF) is used to select CH. Each node constructs an undirected bipartite graph using its neighbor and accessible channel lists and calculates their CHDF value. A node with higher CHDF value is selected as CH. Each cluster comprises a secondary CH to combat the re-clustering issue for mobile nodes. In [9], a cluster formation approach using fault-tolerant backbone construction is introduced for CRN. They implemented a cluster-based back-bone formation protocol to provide virtual backbone without common control channels. Each node sends their node ID and node degree to all its neighbors and compares its node degree with other nodes. The node with the highest node degree amongst neighbors becomes a CH and in case of same node degree, the node with the lowest ID is selected as a CH. Node degree is referred to as the number of neighbor nodes. They also considered dynamic change of network conditions such as the link failure between ordinary node and its CH, departure of CH and departure of GN.

In [10], a multi-channel-based clustering is proposed where the CH is being selected through the node degree. The node with the highest node degree is selected as CH and the closest node to the CH joins as an MN. However, they didn't deeply consider other aspects in cluster formation such as residual energy and channel quality. Energy-efficient Cluster Head Selection (ECHS) protocol is proposed in [11] for CR sensor network, where the nodes that have a larger number of available channels and more residual energy are selected to serve as CHs. To reduce co-channel interference between neighboring clusters, the frequency assignment depends mainly on the coordination between CHs inside a given region and between neighboring regions via message exchange over CCC. To tackle the hot spot problem caused by the high traffic near the sink, ECHS forms many small clusters near the sink to share the forwarding load of the rest of the network. A new Virtual Links Weight-Based Clustering (VLWBC) algorithm for mobile ad-hoc network is presented in [12] to increase network stability. VLWBC not only determines the node's weight using its own features but also considers the direct effect of the feature of adjacent nodes such as link stability and consumed energy by two linked nodes. It determines the weight of virtual links between nodes and the effect of the weights on determining node's final weight and the highest weight is assigned to the best choices for being the CHs. An energy efficient reservation based-based clustering approach has been proposed in [13], where each node knows the time of being a CH and no need to send a message for identification of CH in the network. Therefore, the number of control messages has been decreased and more energy is saved. This method is a round based method consisting of reservation and clustering phase. In [14], a greedy heuristic algorithm is proposed to identify the effect of channel heterogeneity on cluster formation. They considered each member of a cluster should have at least two common channels, one as the main common channel and another act as backup. Each node computed its selection factor using common channel availability matrix and relative channel reward matrix and node with higher selection factor declared itself as CH.

In the most of CRAHN researches, it is assumed that there exists a predetermined CCC. However, in some real environments, it has some problems such as jamming attack and traffic concentration on the dedicated common control channel. In [15], Min-Gyu Kim et al. proposed a cluster-based reliable dynamic common channel setup method without any predetermined channel, in which common channels are dynamically setup based on each node's channel availability. In [16], a graph cut based clustering algorithm for CRAHN is described without considering the predefined CCC. They performed spectrum aware clustering where hoping based discovery protocol is used for neighbor selection. SU similarity matrix is made using two components ratio of common channels (RCC) and relative position similarity (RPS) and then based on the min-cut algorithm they performed clustering. They also considered three-phase transmission period for intra-cluster and inter-cluster communication.

Machine learning in cognitive radios has recently gained a lot of interest in the literature to reduce the complexity and achieve efficiency in real-time spectrum sensing and resource allocation [17]. Among the many different machine learning techniques, reinforcement learning (RL) plays a key role in dynamic and complicated wireless environments. It allows an agent to discover the spectrum situation and take actions using trial and error to maximize the cumulative rewards. Q-learning is the representative reinforcement learning.

In [18], a reinforcement learning-based spectrum-aware clustering algorithm is proposed that allows a member node to learn the energy and cooperative sensing costs for neighboring clusters to achieve an optimal solution. By modeling the network energy consumption in terms of cooperative channel sensing and data communication, the optimal number of clusters is determined. For clustering, it allows that member nodes to learn an optimal policy for choosing optimal clusters based on local decision accuracy and energy consumption. In [19], spectrum-aware cluster-based routing (SMART) algorithm is implied for route selection scheme in CRN utilizing RL. In this method, the node with maximum common channels is selected as a CH. It estimates channel state for the next time instant and uses this estimation to rank and select the operating channel in clustering. In [20], the optimal band and channel selection mechanism for cluster-based CRN is proposed. Based on each MN's reporting in terms of spectrum sensing and traffic demand, CH determines cluster's operating band and channel using Q-learning. Medium access control (MAC) protocols plays a significant role in CRAHN like exploiting spectrum opportunities, scheduling resources, managing PU interference, and coordinating the coexistence of PUs and SUs. In [21], an ample review of MAC protocols in CRAHN is described in details. A maximum edge biclique (MEB) based clustering is discussed in [22], where spectRum Aware cRoss-layEr (RARE) MAC protocol is proposed not only for cluster formation but also a delay-aware routing protocol is proposed for faster data delivery from sender to receiver.

Note that in CRAHN to form and maintain network clusters we need to take into account many considerations as indicated in this section such as energy, member node connectivity, primary statistics, channel quality and coexistence with neighboring clusters. In addition, due to the dynamics of CR environments, interaction with environment as a form of learning is required. In this paper, multi-objective functions using reinforcement learning are taken into account in the clustering design.

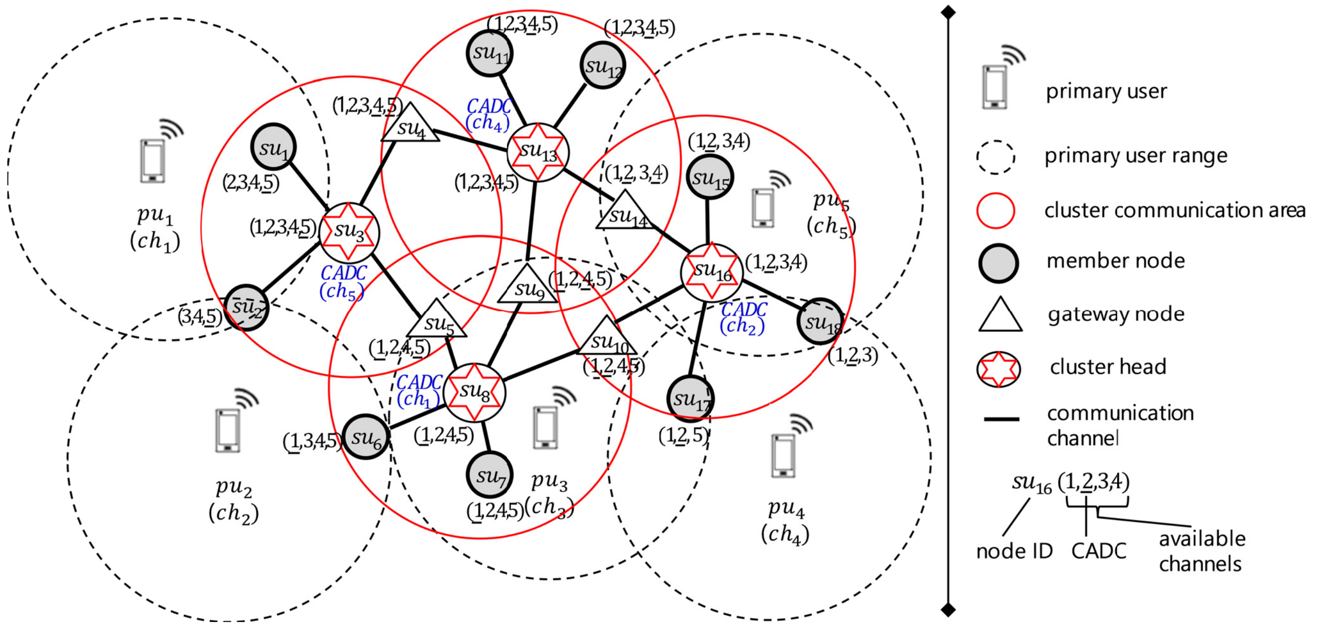

FIGURE 1. Proposed network model.

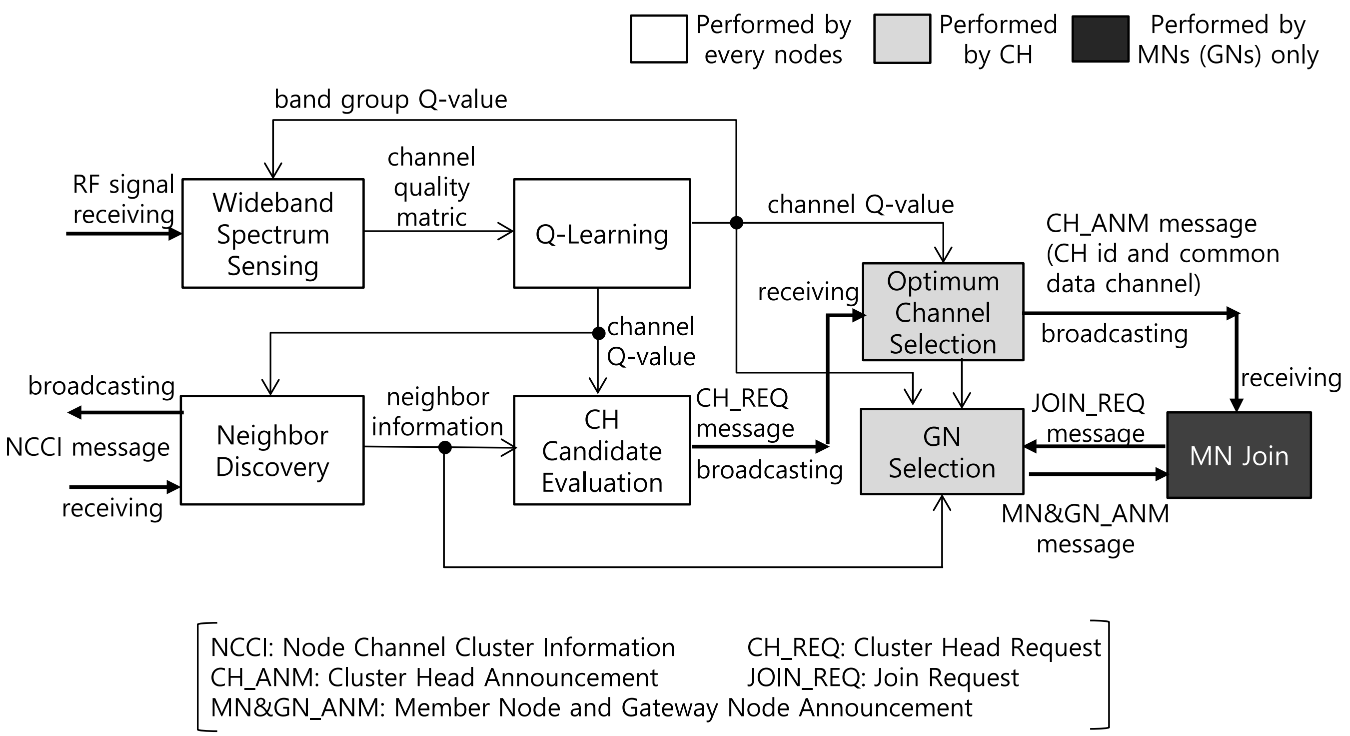

FIGURE 2. Proposed system architecture for CRAHN.

III. PROPOSED SYSTEM ARCHITECTURE

In this paper, we consider distributed cognitive radio ad-hoc networks, in which in the given geographical area there exist P primary users $\boldsymbol{PU} = \{pu_1, pu_2, \cdots, pu_P\}$ and S secondary users $\boldsymbol{SU} = \{su_1, su, \cdots, su_S\}$. Each primary user may occupy one of C channels $\boldsymbol{CH} = \{ch_1, ch_2, \cdots, ch_C\}$. SUs are equipped with a single transceiver and able to sense the spectrum to determine the vacant channels that are not used by near primary users temporary so that the use of the vacant channels by SUs does not cause any harmful interference to PUs. We assume that there exists a predetermined common control channel (CCC) for CRAHN so that SUs are able to exchange some control messages using the CCC to form initial CRAHN clusters or reconfigure the clusters.

Fig. 1 shows the network model considered in this paper. PUs are randomly located in the network area and each one utilizes one of channels. PU operation is modeled as ON(busy)-OFF(idle) process in this paper. The busy and idle periods of any PU channel is assumed to be independent random variables. SUs are assumed to sense the primary signal if they are within the PU range. Nodes that are within the PU range and the PU is in active (busy) should not use the PU's channel. In Fig. 1, node su_1 can't use the channel of pu_1 since it stays within the PU range of pu_1. The actual PU range is determined by the required minimum primary signal power that should be sensed by the SUs. The PU range depends on PU transmission power, sensing requirements for primary protection and wireless channel statistics. For simplicity, in this paper we modeled the PU range as a predefined circle.

As shown in Fig. 1, CRAHN consists of CHs, MNs and GNs. Each cluster has one CH and multiple MNs and GNs. As explained in Section I, the CH is elected by distributed manner using the proposed Q-learning based multi-objective functions and it determines a common active data channel (CADC) and a common backup data channels (CBDC). GNs are connected more than two CRAHN clusters and can relay data from one cluster to other cluster that is known as inter-cluster communication. MNs and GNs of a cluster are one hop neighbors of the CH.

Fig. 2 shows the system architecture of the proposed method. The wideband (band group) spectrum sensing module periodically monitors wideband spectrum and computes channel quality metric for all channels. Based on the historical channel quality metric and expected rewards, the Q-values for channel and band group are updated at the Q-learning module. The neighbor discovery module is used to obtain neighbor node's status, their channel Q-value and reachability information with other nodes and clusters that are broadcasted on the CCC as NCCI (Node Channel Cluster Information) message. Based on the local and neighbor node information, each node determines which node deserves to be a CH and sends the CH_REQ (Cluster Head Request) message to that node, which is done at the CH candidate evaluation module. According to the criteria used in this paper, if a node received enough CH_REQ messages, then it can be a CH and determines the optimum CADC and CBDC. CADC is used for the intra data transmission within the cluster and CBDC is an alternate channel. If the primary system appears on CADC, then the cluster nodes switch to the CBDC for their data communication. The CH broadcast CH_ANM (Cluster Head Announcement) message, in which CH identification and CADC/CBDC are included. At any node that received CH_ANM, if the CADC is available channel to the node, then it replies JOIN_REQ (Join Request) message to the CH. Based on the member node's other cluster reachability and channel Q-value, among the member nodes the optimum GN set for inter-cluster communication is derived by the CH. The CH broadcasts MN&GN_ANM (Member Node and Gateway Node Announcement) message.

Fig. 3 describes the proposed channel operation procedure. In our proposed approach, a CCC is considered network wide initially for cluster formation initiation and reconfigure the cluster. After each cluster formed, CH will select a CADC and CBDC to maintain intra- or inter- cluster communication. Each node in cluster should periodically exchange NCCI messages to report current channel and neighbor conditions. Each node periodically senses the wideband spectrum (procedure (1)) so that it is able to derives all available channels and channel statistics. As we assumed, there exists a CCC to exchange control messages at the initial clustering (procedure (2)) or reformation of the cluster (procedure (7)). Once the cluster is formed, then intra- or inter- cluster communications are performed on the CADC (procedure (3)). Using the CADC, nodes periodically exchange NCCI messages to report current channel and neighbor conditions (procedure (4)). After cluster formation, CH needs to periodically broadcast CH_ANM message on CCC to allow new nodes to join the cluster or the nodes of other clusters to recognize the CH and its CADC (procedure (5)). When primary signal is detected, nodes in the cluster switch to CBDC for their data communication (procedure (6)).

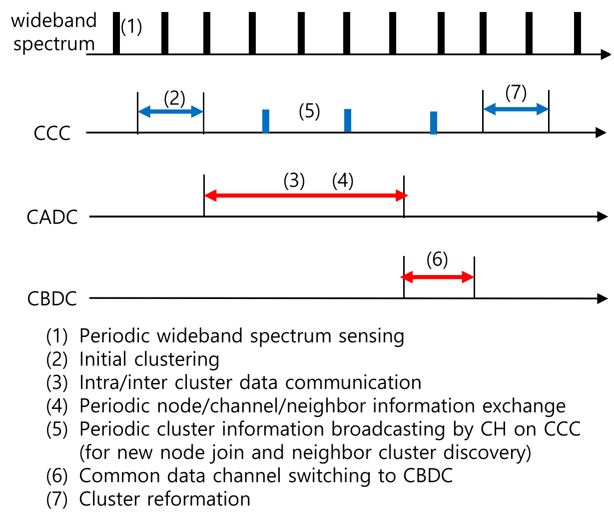

FIGURE 3. Proposed channel operation procedure.

In the proposed system architecture, we have considered the cluster formation in CRAHN that can maximize the proposed multi-objective functions. It should be noted that in this paper we mainly focus on distributed cluster formation but details about time synchronization between nodes, medium access control (MAC) protocol and inter-cluster routing protocol are out of scope of this paper. The contributions of this paper are summarized as follows:

- Learning based channel quality evaluation: We have used Q-learning for measuring the available channel's quality metric. The Q-value is used in several purposes such as band group decision for spectrum sensing, CH selection, CADC/CBDC selection and GN selection.
- Neighbor cluster reachability: In our approach, we have used neighbor cluster reachability information as a factor for CH selection. Since we consider distributed networks, inter-cluster networking is also very important

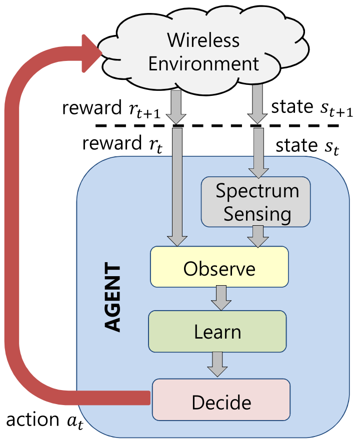

FIGURE 4. Reinforcement learning model.

during cluster formation period. If any node has the reachability towards neighboring cluster then it has a higher probability to be CH as well as a GN.

- Multi-objective CH selection: Since CH is the forerunner in a cluster, it's important to define the CH selection function efficiently. In our algorithm, we have presented the multi-objective functions, in which residual energy, intra-cluster member node connectivity, inter-cluster reachability, available channel list and channel quality are taken into account for CH selection. In this way, the cluster lifetime, node degree, network connectivity, number of common channels and common data channel quality in a cluster can be enhanced.
- Gateway node selection: GNs play a pivotal role in maintaining communication between clusters. In our method, the optimum set of GNs is derived that can minimize redundancy and maximize inter-cluster reachability. While nominating the GN, we used Q-values to provide reliable inter-cluster communication links

IV. Q-LEARNING BASED CLUSTER FORMATION

In this section, the details of the proposed Q-learning based CR cluster formation are explained. First Q-learning for the proposed system architecture to update Q-value is introduced. Then for neighbor discovery, NCCI message format and its properties are presented. Using the channel Q-values and neighbor node information, CH and common data channel selection mechanisms are explained. GNs are determined by GH also based on the channel Q-value and inter-cluster reachability.

A. Q-LEARNING MODEL

Reinforcement learning (RL) methods essentially deal with the solution of optimal control problems using on-line measurements by interacting with an environment. RL is suitable to apply the CRAHN clustering because it can well capture the dynamics of the network topology and spectrum usage. Q-learning is a model-free reinforcement learning algorithm which includes an agent, a set of states $\boldsymbol{S}$, and a set of actions $\boldsymbol{A}$. By performing an action $a \in \boldsymbol{A}$, the agent transitions from state to state. The agent in a state s interacts with the environment with an action a to learn the environment, while depending on outcome it acquires a reward value $r(s, a)$ as shown in Fig. 4.

The goal of the agent is to maximize its total reward. It does this by adding the maximum reward attainable from future states to the reward for achieving its current state, effectively influencing the current action by the potential future reward. This potential reward is a weighted sum of the expected values of the rewards of all future steps starting from the current state.

Suppose at each time t, the agent selects an action a_t, observes a reward r_t, enters a new state s_{t+1}, then Q-value of $Q(s_t, a_t)$ is updated as in (1).

$$Q(s_t, a_t) = (1-\alpha)\, Q(s_t, a_t) + \alpha \left\{ r_t + \gamma \cdot \underbrace{max}_{a} Q(s_{t+1}, a) \right\} \quad (1)$$

where α is the learning rate; γ is the discount factor for the future reward.

B. CHANNEL QUALITY MONITORING AND Q-VALUE UPDATE FOR CR CLUSTERING

As shown in Fig. 2, each node periodically senses spectrum and measures channel quality. In this paper, based on the channel quality metric, the Q-value of (1) is updated. The Q-value is used for optimum band group decision for wideband spectrum sensing and for clustering procedure. Generally, in spectrum sensing, it is impractical to sense the entire spectrum (kHz~GHz) during the given sensing time. Therefore, in our model, the entire operational band is partitioned into G band groups $\boldsymbol{BG} = \{bg_1, bg_2, \cdots, bg_G\}$ and each band group consists of M channels as shown in Fig. 5.

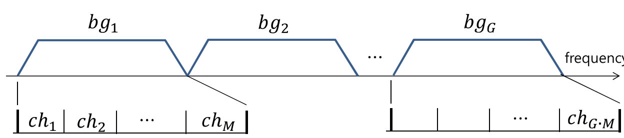

FIGURE 5. Band group structure.

The wideband spectrum sensing is performed for one of the band groups. The band group to be sensed is determined by Q-learning algorithm. The Q-table architecture used in this paper is shown in Fig. 6. The states in the Q-table represent SU node identifications so that each node has single state. The actions represent band groups for spectrum sensing.

The Q-value for band group bg_g of su_s is $Q(su_s, bg_g)$ and it is computed as,

$$Q(su_s, bg_g) = \frac{1}{M} \sum_{c=(g-1)\cdot M+1}^{g\cdot M} Q(su_s, ch_c) \quad (2)$$

where, $Q(su_s, ch_c)$ is the Q-value for channel ch_c of su_s.

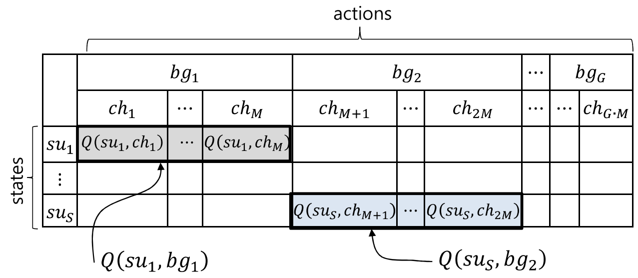

FIGURE 6. Proposed Q-table architecture.

As in (2), $Q\left(su_s, bg_g\right)$ is the average of the Q-values of su_s for the channels of bg_g. The Q-learning module of the secondary user su_s selects the band group bg_g^* for spectrum sensing that has the highest band group Q-value as,

$$bg_g^* = \underset{bg_g}{argmax}\, Q\left(su_s, bg_g\right) \tag{3}$$

Once the band group bg_g^* is determined by the Q-learning module, the spectrum sensing module periodically performs wideband spectrum sensing N_s times consecutively for bg_g^* as shown in Fig. 7. After N_s times, based on the measured channel quality metric, the channel Q-values and band group Q-values are updated. Then the next band group for sensing is selected again. The sensing interval is defined as T_{si} so that the band group section action decision interval can be $N_s \times T_{si}$. The same band group can be reselected consecutively so that the band group sensing time ($T_{bg_g^*}$) for the selected bg_g^*, is the multiple integer times of ($N_s \times T_{si}$).

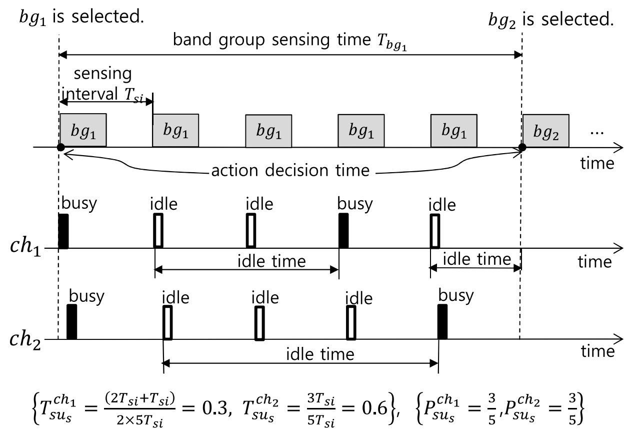

FIGURE 7. Band group sensing and channel quality measurement.

The channel-based Q-values are updated as in (4) when the current band group sensing is finished, i.e., the new different band group is selected for spectrum sensing.

$$Q\left(su_s, ch_c\right) = (1-\alpha)\, Q\left(su_s, ch_c\right) + \alpha \left\{ r_{su_s}^{ch_c} + \gamma \cdot \underbrace{max}_{ch_c} Q\left(su_s, ch_c\right) \right\} \tag{4}$$

The reward $r_{su_s}^{ch_c}$ is the weighted sum of the normalized average idle time and the estimated idle time probability for each channel during the channel observation time as in (5).

$$r_{su_s}^{ch_c} = \omega_1 \cdot T_{su_s}^{ch_c} + \omega_2 \cdot P_{su_s}^{ch_c} \tag{5}$$

$$T_{su_s}^{ch_c} = \frac{average\ ch_c\ channel\ idel\ time}{band\ group\ bg_g^*\ sensing\ time} = \frac{E\left[T_{idle}^{ch_c}\right]}{T_{bg_g^*}} \tag{6}$$

$$P_{su_s}^{ch_c} = \frac{number\ of\ idle\ observations\ for\ ch_c}{number\ of\ ch_c\ sensing\ trials\ during\ T_{bg_g^*}} \tag{7}$$

where, $\omega_1 + \omega_2 = 1$; $T_{su_s}^{ch_c}$ and $P_{su_s}^{ch_c}$ are the normalized average idle time and estimated idle time probability of ch_c for su_s. Busy or idle decision can be made using any sensing algorithms. In the case of energy detection-based spectrum sensing, if the energy received signal is greater than a threshold then it indicates the channel is occupied by a primary user.

For the example case in Fig. 7, for the secondary user su_s, band group bg_1 was selected for spectrum sensing and after $N_s(= 5)$ sensing intervals, band group bg_2 is selected. For ch_1 and ch_2 of bg_1, even though the estimated idle time probabilities are the same (3 idle times among 5 sensing trials), the normalized average idle times are different. ch_2 shows longer normalized average idle time than that of ch_1. In the proposed cluster formation procedures, channels quality Q-values play a vital role in selecting the CH as well as selecting the common data channels and gateway nodes.

C. NEIGHBOR DISCOVERY AND NODE-CLUSTER-CHANNEL INFORMATION EXCHANGE

In the proposed system, for initial cluster formation, each node performs neighbor discovery procedure. Each node builds a NCCI (Node Channel Cluster Information) message and broadcasts it to its one hop neighbors through CCC, in which NCCI consists of the node property, channel quality Q-values, neighbor node information and reachable neighbor cluster information. The NCCI message format is shown in Fig. 8.

Node property field includes node ID (id) and the current residual energy level $\left(E_{id}^R\right)$. Channel quality field contains the available channel list ($\boldsymbol{AC}_{id} = \{ac_{id}\}$) and the updated Q-value set $\boldsymbol{Q}_{id} = \{Q\left(id, ac_{id}\right)\}$ for all available channels that are measured by spectrum sensing. Neighbor node connectivity field has the neighbor node list ($\boldsymbol{NN}_{id} = \{id\}$) and each neighbor node's available channel list that are obtained by receiving the broadcasted NCCI messages from one hop neighbors. Neighbor cluster reachability field consists of the set of connectable neighbor clusters ($\boldsymbol{NC}_{id} = \{id\}$) and their current CADCs ($\boldsymbol{NCDC}_{id} = \{dc_{id}\}$). The cluster ID uses the CH ID of the cluster. The neighbor cluster information can be acquired by receiving CH_ANM messages that are periodically broadcasted on CCC by neighbor CHs.

For the example case of Fig. 8, the NCCI message is broadcasted by node A and it has four available channels. Node A has four neighbor nodes $\{B, D, E, F\}$ and it is also connectable two neighbor clusters $\{K, L\}$ (i.e., node A is

Node Property	Node ID (id)	Residual Energy (E_{id}^R)
	A	3J
Channel Quality	Available Channel List $\boldsymbol{AC}_{id} = \{ac_{id}\}$	Q-values of the Available Channels $\boldsymbol{Q}_{id} = \{Q(id, ac_{id})\}$
	$\{1,3,4,5\}$	$\{Q(A,1), Q(A,3), Q(A,4), Q(A,5)\}$
Neighbor Node Connectivity	Neighbor Node List $\boldsymbol{NN}_{id} = \{id\}$	Neighbor Node's Available Channel List $\boldsymbol{NAC}_{id} = \{\{\boldsymbol{AC}_{id}\}\}$
	$\{B,D,E,F\}$	$\{\boldsymbol{AC_B}, \boldsymbol{AC_D}, \boldsymbol{AC_E}, \boldsymbol{AC_F}\} = \left\{\begin{matrix}\{1,2,3,4\},\{1,2,4\},\\ \{2,3,4,5\},\{1,2,3,4,5\}\end{matrix}\right\}$
Neighbor Cluster Reachability	Set of Connectable Neighbor Clusters $\boldsymbol{NC}_{id} = \{id\}$	CADCs of the Connectable Neighbor Clusters $\boldsymbol{NCDC_{id}} = \{dc_{id}\}$
	$\{K,L\}$	$\{dc_K, dc_L\} = \{3,4\}$

FIGURE 8. NCCI message format.

already a MN of the neighbor cluster or it can join the cluster using the cluster's CADC). The CADCs of the neighbor cluster K and L are channel 3 and channel 4, respectively. To avoid possible interference between neighbor clusters, the CADCs used by neighbor clusters should not be used by the newly forming cluster. Therefore, the new cluster that covers node A should not use channels $\{3, 4\}$ that are currently used by neighbor clusters $\{K, L\}$. Once the cluster has been formed, all member nodes periodically update and broadcast the NCCI on its CADC.

D. MULTI-OBJECTIVE CLUSTER HEAD FITNESS FUNCTION

By exchanging NCCI messages, each node can have all required neighbor information. In the proposed distributed CH selection procedure, each node evaluates the CH fitness for all neighbor nodes and itself. CH fitness value for a node indicates that how appropriate is the node to be a CH. We propose a multi-objective fitness function to compute the CH fitness value, in which we considered each node's residual energy, channel quality, the number of available channels, the number of connectable neighbor nodes and the number of reachable clusters.

For node j, the channel fitness value CF_j is computed as follows:

$$CF_j = \sum\nolimits_{k\in \boldsymbol{EAC}_j} \left(Q(j,k) \cdot \left|\boldsymbol{N}_j^k\right|\right) \tag{8}$$

$$\boldsymbol{EAC}_j = \boldsymbol{AC}_j - \bigcup\nolimits_{n\in(\boldsymbol{NN}_j \cup j)} \boldsymbol{NCDC}_n \tag{9}$$

$$\boldsymbol{N}_j^k = \left\{n \in \boldsymbol{NN}_j | k \in \boldsymbol{AC}_n\right\} \tag{10}$$

where, $\boldsymbol{EAC}_j$ is the set of effective available channels (EACs) for node j; $\boldsymbol{N}_j^k$ is the set of neighbor nodes that can be connected with node j using the channel k; $|S|$ cardinality operator is a measure of the number of elements of the set S. The set of effective channels of node j is computed such that from the available channel set of node j, the CADCs of neighbor clusters are removed. As in (8), the channel fitness value is the higher for the larger number of effective available channels, the higher Q-values for the EACs and the more number of connectable neighbor nodes using the EACs.

The CH fitness value of node j is defined V_j as in (11).

$$V_j = \beta_1 \frac{E_j^R}{E_{max}} + \beta_2 \frac{CF_j}{CF_{max}} + \beta_2 \frac{RNC_j}{NRC_{max}} + \beta_4 \frac{\left|\boldsymbol{NN}_j\right|}{NN_{max}} \tag{11}$$

$$RNC_j = \left|\boldsymbol{NC}_j \bigcup \left(\bigcup\nolimits_{n\in \boldsymbol{NN}_j} \boldsymbol{NC}_n\right)\right| \tag{12}$$

where, $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$; E_j^R is the residual energy of node j; CF_j is the channel fitness value of node j; RNC_j is the number of reachable neighbor clusters through node j itself and node j's neighbor nodes; $\left|\boldsymbol{NN}_j\right|$ is the number of neighbor nodes of node j. E_{max}, CF_{max}, NRC_{max} and NN_{max} are the predetermined maximum values for normalization.

E. CH AND COMMON DATA CHANNEL SELECTION PROCEDURE

Each node i computes CH fitness values for its all one hop neighbors and node i itself using Eq. (11). Then node i selects the node that has the highest CH fitness value and it is considered as a CH candidate for node i (CH_i^c) as in (13). If node i's V_i is higher than those of any other its neighbor nodes, then node i's CH candidate can be node i itself.

$$CH_i^c = \underset{j\epsilon(\boldsymbol{NN}_i \cup i)}{argmax}\ V_j \tag{13}$$

After selecting a CH candidate node, each node sends CH_REQ message to its CH_i^c node using CCC. If any node j received CH_REQ messages, including self-requesting, more than predetermined ratio as in (14), then node j can be a CH for its neighbor nodes.

$$n_j^{CH\_REQ} \geq \eta \cdot \left(\left|\boldsymbol{NN}_j\right| + 1\right) \tag{14}$$

where, $n_j^{CH\_REQ}$ is the total received CH_REQ messages including self-requesting; η is the percentile threshold.

Fig. 9 shows an example case for the proposed CH selection procedure, in which we assume that there exists no neighbor cluster. After several NCCI message exchanges, each node has the neighbor information as in Fig. 9. Using Eq. (8) and (11), each node computes channel fitness value CF_j and CH fitness value V_j for its neighbor nodes and the node itself. In this example case, V_C of node C is the highest so that every node transmits CH_REQ message to node C and node C will be the CH for node A, B and D. In this example case, we set $\eta = 0.5$; $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.25$; $E_{max} = CF_{max} = NRC_{max} = NN_{max} = 5$.

Fig. 10 shows the cluster formation example, in which multiple clusters are generated simultaneously. As the same way in Fig. 9, each node computes CH fitness value for its neighbor nodes and itself and sends CH_REQ message to the node that has the highest CH fitness value. Node A, B, D and K received at least one CH_REQ message. In this example, we used $\eta = 0.5$ so that node A cannot satisfy the requirement of Eq. (14). As a result, node B, D and K

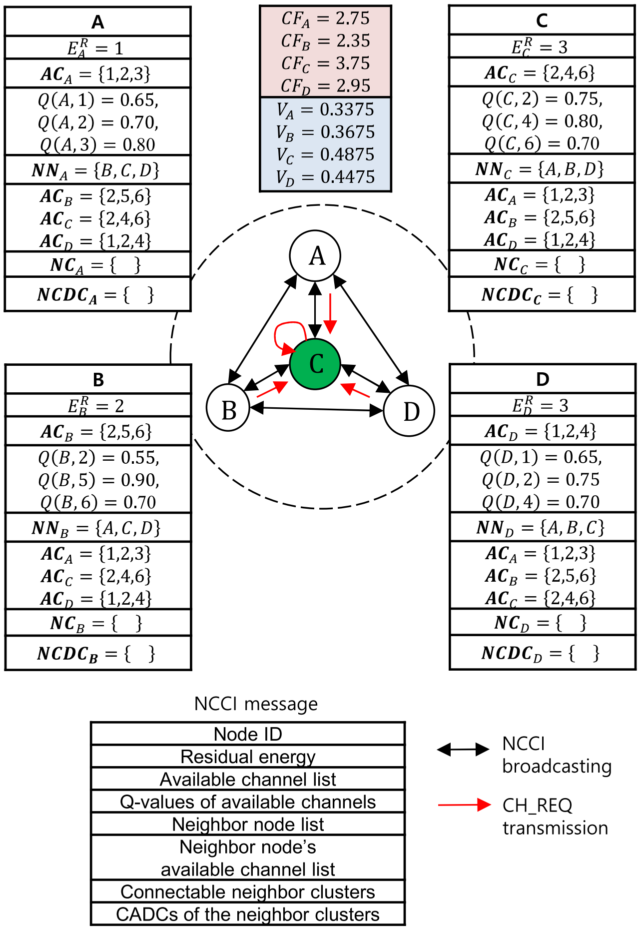

FIGURE 9. Proposed CH selection example: single cluster case.

become CHs and configure their clusters. Suppose node E's available channels include node D cluster's CADC and node K's CADC. Then, node E will be the gateway node for inter-cluster communication.

The CH node j needs to determine the common active data channel for the cluster. To avoid the possible interference between neighbor clusters, the CH should avoid using the CADCs of the neighbor clusters. In this paper, for CADC selection, we considered channel Q-value and possible number of connectable neighbor nodes using the channel. The CADC selection rule for CH j is as follow:

$$CADC_j = \underset{k \in \boldsymbol{EAC}_j}{argmax} \left(Q(j,k) \cdot \left| \boldsymbol{N}_j^k \right| \right) \tag{15}$$

where, $\boldsymbol{EAC}_j = \boldsymbol{AC}_j - \bigcup_{n \in (\boldsymbol{NN}_j \bigcup j)} \boldsymbol{NCDC}_n$ and $\boldsymbol{N}_j^k = \{n \in \boldsymbol{NN}_j | k \in \boldsymbol{AC}_n\}$. The channel among $\boldsymbol{EAC}_j$ having the second largest $Q(j,k) \cdot \left|\boldsymbol{N}_j^k\right|$, if exists, is assigned as the common backup data channel (CBDC), which will replace the current CADC when primary signal is detected on the CADC.

CH j then broadcast CH_ANM message to its one hop neighbors using CCC, in which it contains CH node *id*, CADC and CBDC.

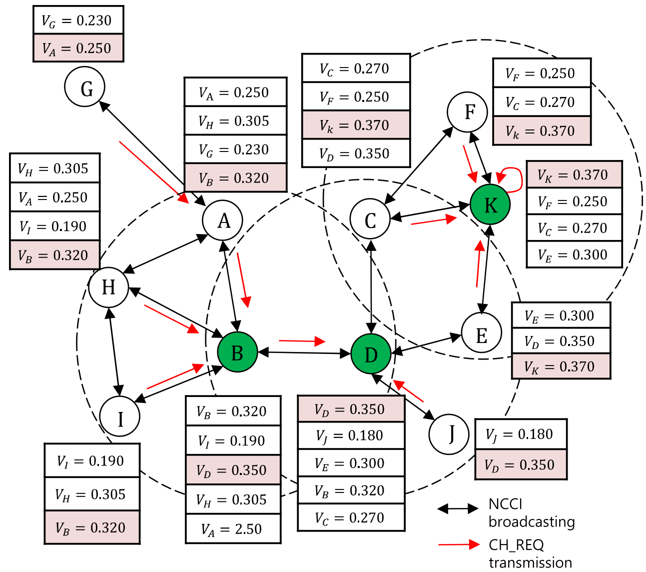

FIGURE 10. Proposed CH selection example: multiples cluster case.

F. MEMBER NODE JOIN AND GATEWAY NODE SELECTION

If a node receives a CH_ANM message and the CADC included in the message is one of the available channels of the node, then the node can join the cluster by sending JOIN_REQ message on the CCC. If a node i received multiple CH_ANM messages from different CHs, then the node will select the optimum CH j that has the largest CH fitness value as in (16).

$$CH_i^* = \underset{j \epsilon (\boldsymbol{CHA}_i)}{argmax}\ V_j \tag{16}$$

where, $\boldsymbol{CHA}_i$ is the set of neighbor nodes that have transmitted CH_ANM message received by node i.

After broadcasting CH_ANM message, CH j receives JOIN_REQ messages from its neighbor nodes and the nodes that sent JOIN_REQ message are assigned as member nodes (MNs). Since CH already has the neighbor cluster reachability information for all neighbor nodes, it knows that which neighbor clusters exist, what CADCs are used by the neighbor clusters and which join requesting nodes can be reachable to which neighbor clusters. This can be derived using the received NCCI messages during the neighbor discovery procedure. If a MN is the only node that can interconnect certain neighbor cluster l, then it is assigned as a gateway node (GN_{j-l}) for neighbor cluster l. If there are multiple member nodes that are able to interconnect with cluster l, then the CH j selects one gateway node (GN_{j-l}) that has the largest average Q-value for $CADC_j$ and $CADC_l$ as in (17).

$$GN_{j-l} = \underset{m | CADC_j, CADC_l \in \boldsymbol{AC}_m}{argmax} \left(\frac{Q(m, CADC_j) + Q(m, CADC_l)}{2} \right) \tag{17}$$

After selecting GNs, the CH broadcasts MN&GN_ANM message, which includes MN *id*, GN *id* and neighbor cluster *id* for each GN.

V. SIMULATION RESULTS

The proposed Q-learning based clustering algorithm has been evaluated using MATLAB. We have randomly deployed secondary nodes in the simulation area of 100 × 100 m^2. The distribution of the initial energy of each secondary user node is uniform in [0.7J, 1.5J]. The simulation parameters used in our system are given in Table I. We have chosen three well-known cluster formation methods to evaluate the performance of our proposed scheme. K-means clustering [6] for cognitive radio condition, which is one of the simplest and popular unsupervised machine learning algorithms and multi-channel-based clustering (MCBC) [10], where the cluster head is being decided based on node degree that can communicate using the common available channels. The third method is maximum edge biclique (MEB) based clustering approach [23] based on Spectrum availability, node speed, power level on each node. In [23], both static and mobile nodes are considered for cluster formation. Since we have only considered static nodes, to compare with this clustering approach we have chosen the node velocity is always 1 for all nodes.

Fig. 11 is one of the generated clustering topologies derived from our proposed algorithm when the transmission range of the nodes is 30m. PUs and secondary nodes are distributed randomly and CH, MNs, GNs as well as CADCs are selected based on the proposed approach. In our proposed distributed network-based algorithm, node without having neighbor cluster information will initiate cluster formation procedure.

TABLE 1. Simulation parameters

Parameters	Value	Parameters	Value
Simulation area	100×100m^2	β_1, β_2, β_3, β_4	0.066~0.8
SU residual energy	0.7~1.5 J	E_{max}	1.5J
Number of SUs	10~40	NN_{max}	10
Number of PUs	4~12	CF_{max}	20
SU transmission range	25~45m	NRC_{max}	10
PU range	50 m	N_s	15
Learning rate, α	0.5	G	2
Discount factor, γ	0.5	M	6
Percentile threshold, η	0.5	Primary $E[ON], E[OFF]$	10 ~30 T_{si}
ω_1, ω_2	0.5		

From Fig. 12 to Fig. 15, we show the clustering performance of the proposed method for various conditions. Except the cases that we explicitly change some parameters, the number of PUs is 12; SU transmission rate is 30m; $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.25$.

As shown in Fig. 12, it is observed that as the transmission range of nodes is increasing, the number of member nodes in each cluster is also increasing. This is because as

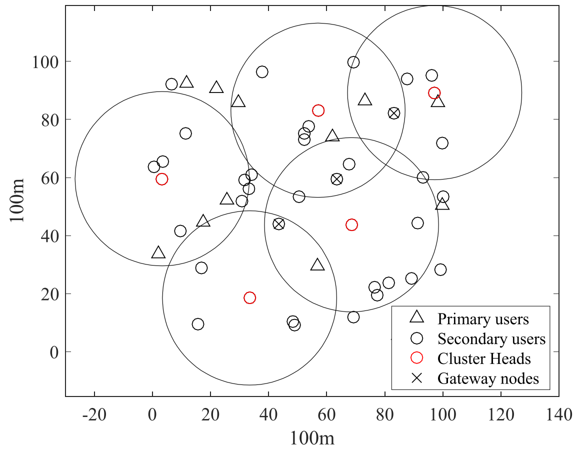

FIGURE 11. Simulation topology of the proposed approach.

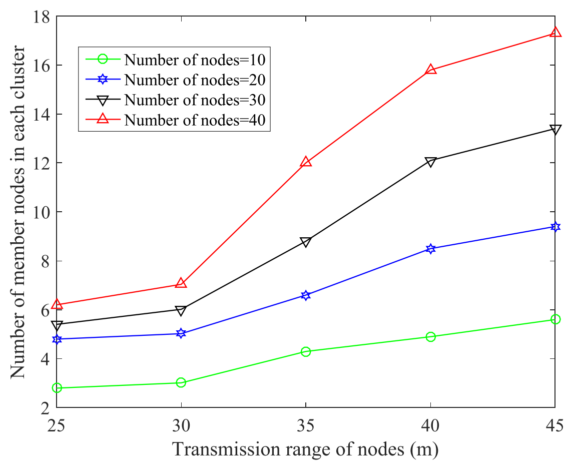

FIGURE 12. Transmission range vs number of member node in a cluster.

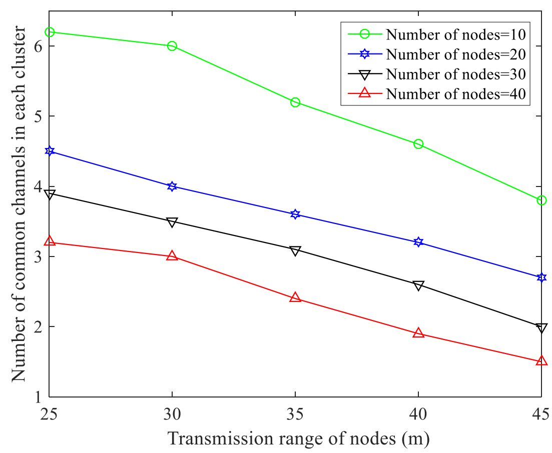

FIGURE 13. Transmission range vs number of common channels in a cluster.

the transmission range of SUs is extending the number of neighbors of a CH is also increasing.

Fig. 13 represents the number of common channels in each cluster with respect to the SU node transmission range. When the PU range is fixed to 50m, as the transmission range of SU node increases, the number of common channels in a cluster

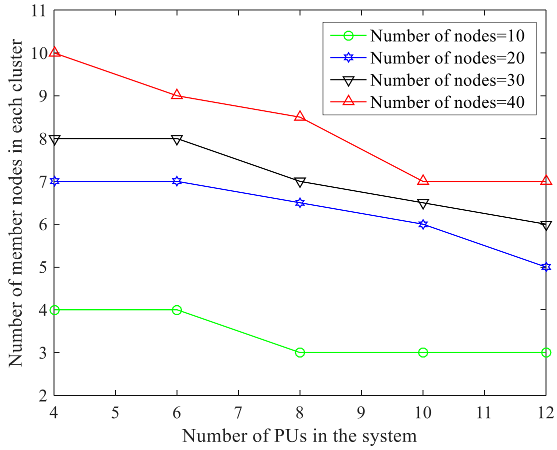

FIGURE 14. Number of PUs vs number of member nodes in a cluster.

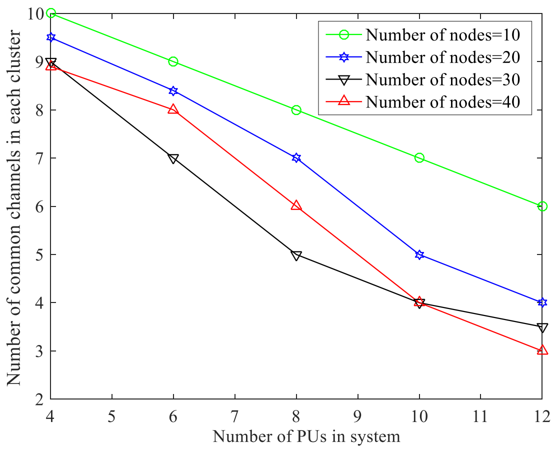

FIGURE 15. Number of PUs vs number of common channels in a cluster.

decreases. If we increase the SU transmission range, then some of one hop neighbors of a CH can be located in the PU range so that those PU's channels are not included in the common channels. Also, we can observe that the more number of nodes in the network results in the smaller number of common channels in a cluster because it makes the higher possibility that some SU nodes are within PU ranges.

Fig. 14 shows the number of member nodes with respect to the number of PUs in the system. The number of MNs in a cluster is decreasing with the increase in the number of primary users in the system. When the number of PUs are increasing the number of available channels for SUs are decreasing which results less number of common idle channels within neighbors of SUs. That's why when the number of PUs are increasing the cluster size is decreasing i.e., member nodes in a cluster is decreasing even if nodes are within the range of the CH. In our simulation, we have used 12 primary channels whereas the number of PUs are varied from 4 to 12. As shown in Fig. 15, the number of common channels is decreasing with the growth of PUs in the system. When all the channels are assigned to PUs, the cluster size is the lowest as well as the number of common channels within a cluster.

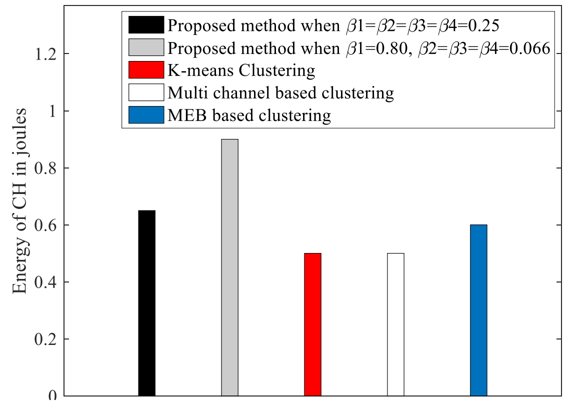

FIGURE 16. Effect of node energy in CH selection.

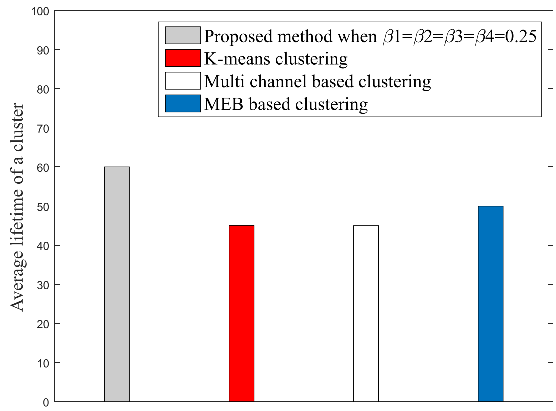

FIGURE 17. Comparison of lifetime of a cluster.

Fig. 16 shows the average energy level of the selected CHs. As we can see the proposed method selects CHs that have more residual energy compared with those of other conventional methods. When we set all β parameters for fitness value of CH of Eq. (11) be 0.25 ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.25$), the average residual energy of the selected CHs is about 0.65J which is 35% higher than other compared methods. When we set $\beta_1 = 0.8$, $\beta_2 = \beta_3 = \beta_4 = 0.066$, which means that we give more weight for the energy factor in Eq. (11) than those for other factors, the average residual energy of CHs is about 0.91J and it is almost 90% energy performance increasing compared with others. In MEB based clustering, the node energy is being normalized for which reason CH election almost depends on node degree and common idle channel but it performs better than the other two methods.

In Fig. 17, we observe that the lifetime of the cluster in our proposed method is much higher than those of two compared methods. This is because we have formulated each process of cluster formation in such a way that enhances the lifetime of clusters. The constructed cluster can be broken or should be reconfigured when the CH does not have enough energy, the CADC (or CBDC) is not any more available. In the proposed method, we utilize Q-learning based channel evaluation model and we have chosen the CH which have higher CH fitness value as well as gateway node which have the largest average Q-value of CADCs in connecting CHs. As we

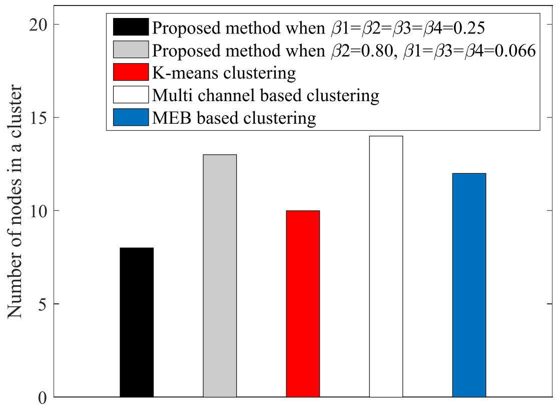

FIGURE 18. Comparison of proposed method with existing methods in terms of node degree.

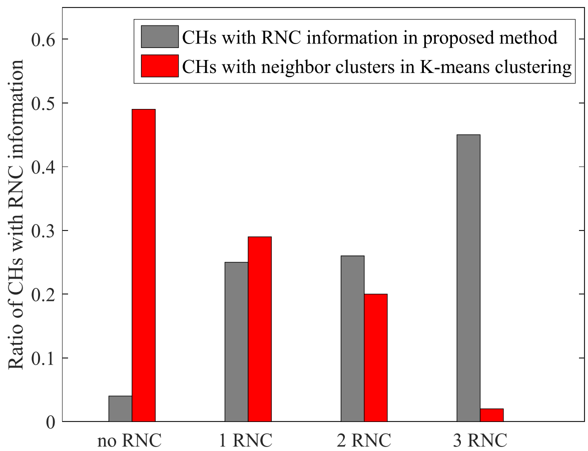

FIGURE 19. Number of reachable neighbor clusters.

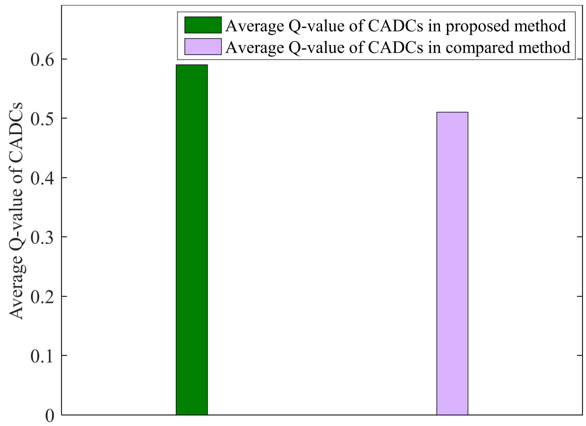

FIGURE 20. Comparison of average Q-value of CADCs.

can see in Fig. 17, the proposed method can have 30% longer cluster life time compared with conventional methods. MEB based clustering performs better than K-means and MCBC only because it has more common idle channel within the cluster but it doesn't perform better than our proposed method since we've considered CADC (as well as CBDC) as with highest Q-valued channel.

In Fig. 18, the number of member nodes of a cluster is compared. In K-means clustering, the node that is closest to the centroid of the cluster is selected as CH and in multi-channel-based clustering the node that has the largest number of neighbor nodes that can share the common available channels is selected as CH. Since Cluster size should be higher in MEB based clustering and MCBC is node degree-based clustering, both methods outperform our method when all variables in CH selection is considered with equal impact factor. But when we increase the beta value for node energy then our method performs better than MEB based clustering is quite competitive with the multi-channel-based clustering.

In Fig. 19 and Fig. 20, we evaluated the performance of inter-cluster communication, in which neighbor cluster reachability and selected gateway node channel quality are compared. For this simulation, we used 10 different topologies where for each topology the transmission range of SUs is $25m \sim 45m$. We have 40 SU nodes and 12 PUs. Fig. 19 shows the ratio of CHs having the given number of reachable neighbor clusters. Although in K-means algorithm they didn't consider the neighbor reachability information, we have assumed that neighbor cluster information is delivered to CH by the member nodes even in K-means algorithm. After a cluster is formed, we have checked the number of reachable clusters for each CH. In the proposed method, the reachable number of clusters is also considered in the CH fitness function. For the proposed method, only 5% of CHs did not have any neighbor cluster, but about 50% of CHs have three reachable neighbor clusters. On the other hands, about 50% of CHs in K-means clustering did not have any reachable neighbor clusters and only 2% of CHs were able to have three reachable neighbor clusters. In this paper, we propose a Q-learning based cluster formation method for distributed CRAHN. As a reinforcement learning, Q-learning is used to evaluate and estimate channel quality at each node in terms of opportunistic channel access possibility and sustainable operation time. The Q-values for the available channels at each node are exchanged between neighbor nodes and then the proposed CH selection and common active data channel (CADC) decision processes are applied. We also propose a GN selection method that can provide effective inter-cluster connectivity. In the proposed cluster formation for CRAHN, we have defined multi-objective function that considers channel quality, network life time, even energy consumption, number of member nodes, network connectivity and coexistence with neighbor networks

VI. CONCLUSION

In this paper, we proposed a Q-learning based clustering mechanism for cognitive radio ad-hoc networks. For initial wideband spectrum sensing, the dynamic Q-table update procedure for each channel and band group has been defined to select a band group to be sensed. The Q-value of a channel captures primary user's activity and their operation patterns and it is used to evaluate channel quality at each secondary user. One of the prime concerns in cluster formation in how to construct a stable cluster despite the influence of PU activity.

For CR ad-hoc cluster formation, multi-objective CH selection functions have been defined, in which the node residual energy, channel quality fitness value using Q-learning, neighbor cluster reachability and number of neighbor nodes having common available channels are considered. The neighbor cluster reachability information helps each node to reduce not only the inter-cluster interference but also ensure effective communication with neighbor clusters. For CADC and GN selection, we utilized Q-values of channels to provide reliable and stable intra and inter cluster communication service.

For the simulation study, we implemented CR ad-hoc network simulator using MATLAB that could set various CR network topologies and network conditions. For the proposed clustering mechanism, we discussed the effects of various network conditions on the number of member nodes and the number of common available channels of clusters. Comparing with other conventional clustering schemes, it is evident that the performance of our method outperforms in terms of CH residual energy, cluster lifetime, neighbor cluster reachability and channel quality of the selected GNs. As a further research, we will focus on the optimal routing path selection through inter cluster communication using some machine learning techniques such as reinforcement learning, fuzzy logic, or neural network.

REFERENCES

[1] *Spectrum Policy Tasks Force Report*, Federal Commun. Commission, Washington, DC, USA, Nov. 2002.

[2] J. Mitola and G. Q. Maguire, Jr., "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Apr. 1999.

[3] T. Chen, H. Zhang, G. M. Maggio, and I. Chlamtac, "CogMesh: A cluster-based cognitive radio network," in *Proc. IEEE DySPAN*, Apr. 2007, pp. 168–178.

[4] A. C. Talay and D. T. Altilar, "United nodes: Cluster-based routing protocol for mobile cognitive radio networks," *IET Commun.*, vol. 5, no. 15, pp. 2097–2105, Oct. 2011.

[5] S. Kumar and A. K. A. Singh, "A localized algorithm for clustering in cognitive radio networks," *J. King Saud Univ.-Comput. Inf. Sci.*, pp. 1–8, 2018.

[6] B. Benmammar, M. Taleb, and F. Krief, "Diffusing-CRN k-means: An improved k-means clustering algorithm applied in cognitive radio ad hoc networks," *Wireless Netw.*, vol. 23, no. 6, pp. 1849–1861, 2016.

[7] R. M. Eletreby, H. M. Elsayed, and M. M. Khairy, "CogLEACH: A spectrum aware clustering protocol for cognitive radio sensor networks," in *Proc. 9th Int. Conf. Cognit. Radio Oriented Wireless Netw. Commun. (CROWNCOM)*, 2014, pp. 179–184.

[8] N. Mansoor, A. Islam, M. Zareei, S. Baharun, and S. Komaki, "Construction of a robust clustering algorithm for cognitive radio ad-hoc network," in *Cognitive Radio Oriented Wireless Networks* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 156. 2015, pp. 759–766.

[9] S. Kumar and A. K. Singh, "Fault tolerant backbone construction in cognitive radio networks," *AEU-Int. J. Electron. Commun.*, vol. 87, pp. 76–86, Apr. 2018.

[10] R. K. Berwer and S. Kumar, "Multi channel-based clustering in cognitive radio networks," in *Proc. Int. Conf. Smart Technol. Smart Nation (SmartTechCon)*, 2017, pp. 665–670.

[11] E. Helal, A. Khattab, and Y. A. Fahmy, "Energy-efficient cluster head selection for cognitive radio sensor networks," in *Proc. 28th Int. Conf. Microelectron. (ICM)*, 2016, pp. 205–208.

[12] A. Karimi, A. Afsharfarnia, F. Zarafshan, and S. A. R. Al-Haddad, "A novel clustering algorithm for mobile ad hoc networks based on determination of virtual links' weight to increase network stability," *Sci. World J.*, vol. 2014, 2014, Art. no. 432952.

[13] A. Zahedi, M. Arghavani, A. Arghavani, and F. Parandin, "Energy efficient reservation-based cluster head selection in WSNs," *Wireless Pers. Commun.*, vol. 100, no. 3, pp. 667–679, Jun. 2018.

[14] M. M. A. Osman, S. K. S. Yusof, and N. N. N. A. Malik, "Impact of channel heterogeneity on clustering formation in cognitive ad hoc radio networks," *Wireless Pers. Commun.*, vol. 96, no. 3, pp. 4613–4627, 2017.

[15] M.-G. Kim, S.-J. Yoo, Y.-U. Jang, and K.-E. Lee, "Reliable cluster-based common channel setup method for DSA CR networks," in *Proc. 5th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, 2013, pp. 155–160.

[16] H. Zhang, N. Xu, F. Xu, and Z. Wang, "Graph cut based clustering for cognitive radio ad hoc networks without common control channels," *Wireless Netw.*, vol. 24, no. 1, pp. 209–221, 2018.

[17] N. Abbas, Y. Nasser, and K. E. Ahmad, "Recent advances on artificial intelligence and learning techniques in cognitive radio networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, pp. 1–20, 2015.

[18] I. Mustapha, B. M. Ali, A. M. F. Rasid, A. Sali, and H. Mohamad, "An energy-efficient spectrum-aware reinforcement learning-based clustering algorithm for cognitive radio sensor networks," *Sensors*, vol. 2015, no. 15, pp. 19783–19818, 2015.

[19] Y. Saleem, K.-L. A. Yau, H. Mohamad, N. Ramli, M. H. Rehmani, and Q. Ni, "Clustering and reinforcement-learning-based routing for cognitive radio networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 146–151, Aug. 2017.

[20] S.-J. Jang, C.-H. Han, K.-E. Lee, and S.-J. Yoo, "Reinforcement learning-based dynamic band and channel selection in cognitive radio ad-hoc networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, pp. 1–25, 2019.

[21] M. Zareei, A. K. M. M. Islam, S. Baharun, C. Vargas-Rosales, L. Azpilicueta, and N. Mansoor, "Medium access control protocols for cognitive radio ad hoc networks: A survey," *Sensors*, vol. 17, p. 2136, 2017.

[22] 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.

[23] M. Zareei, E. M. Mohamed, M. H. Anisi, C. V. Rosales, K. Tsukamoto, and M. K. Khan, "On-demand hybrid routing for cognitive radio ad-hoc network," *IEEE Access*, vol. 4, no. 1, pp. 8294–8302, 2016.

MD ARMAN HOSSEN received the B.Sc. degree in electronics and telecommunication engineering from the Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh, in 2014. He is currently pursuing the M.S. degree with the Graduate School of Information and Communication Engineering, Inha University, Incheon, South Korea. From 2016 to 2018, he worked as an Engineer with the Fiber@Home Ltd., one of the largest Telecom Company in Bangladesh. His current research interests include cognitive radio networks and machine learning applications in wireless sensor networks.

SANG-JO YOO received the B.S. degree in electronic communication engineering from Hanyang University, Seoul, South Korea, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, in 1990 and 2000, respectively. From 1990 to 2001, he was a member of Technical Staff with the Korea Telecom Research and Development Group, where he was involved in communication protocol conformance testing and network design fields. From 1994 to 1995 and 2007 to 2008, he was a Guest Researcher with the National Institute Standards and Technology, USA. Since 2001, he has been with Inha University, where he is currently a Professor with the Department of Information and Communication Engineering. His current research interests include cognitive radio network protocols, ad hoc wireless networks, MAC and routing protocol design, wireless network QoS, and wireless sensor networks.