A Hierarchical Routing Protocol based on Energy Consumption Weight Clustering Scheme for Cognitive Radio Networks

Yihang Du[1, a], Hu Jin[1], Lijia Wang[2] and Lei Xue[1]
[1]*Electronic Countermeasure Institute, National University of Defense Technology, Hefei, China*
[2]*CTTL-Terminals of China Academy, Telecommunication Research of MIIT, Beijing, China*

Keywords: Hierarchical routing protocol, clustering, energy consumption ratio, multi-agent Q-learning.

Abstract: In order to reduce the network congestion and data forwarding times, a hierarchical routing protocol based on energy consumption weight clustering scheme is proposed. Firstly, the concept of Energy Consumption Weight (ECW) is introduced. Then the clustering problem is modelled as a complete bipartite graph decomposition problem with maximum weights and a greedy clustering scheme based on ECW is presented to minimize the energy consumption of intra-cluster transmission. Subsequently, the Equal Reward Timeslots based Conjectural Multi-Agent Q-Learning (ERT-CMAQL) is applied to optimize routing and resource allocation in inter-cluster communication. Simulation results show that the proposed hierarchical routing scheme outperforms the flat routing protocol in terms of system energy consumption and packet transmission latency, and it effectively reduces the number of nodes involved in operation and decision-making in the multi-agent learning scheme when the size of network is large.

1 INTRODUCTION

With the explosive growth of the communication divices, the scarcity of spectrum resource is becoming a severe problem. Cognitive Radio (CR) is a promising technology to break the static channel assignment policy and realize the Dynamic Spectrum Access (DSA). It enables the Secondary Users (SUs) to opportunistically access the spectrum that is not occupied by the licensed users for enhancing the spectrum utilization rate (Chen et al. 2016). In order to expand the deployment scope and ensure the flexibility as well as robustness of the system, Cognitive Radio Network (CRN) usually adopts the distributed network architecture, and transmits data from the source node to the destination node in multi-hop manner. Therefore, it is of significant importance to take routing into consideration in CRN (Cesana et al. 2011).

Reinforcement Learning (RL) has been widely used in the routing design in CRN since it does not require the priori information such as spectrum statistics and network topology. Furthermore, RL methods have good adaptability to the dynamic and unpredictable nature of CRN. A cross-layer routing protocol based on Prioritized Memories Deep Q-Network (PM-DQN) was proposed in (Du et al. 2018) to solved the joint routing and resource management problem in the large-scale CRN. However, it applied single- agent learning framework, which had low convergence speed and high signaling overhead. A conjecture-based multi-agent Q-learning scheme was presented in (Cao et al. 2014) to perform route selection in a partially observable environment. The routing problem was formulated as a stochastic game (SG) in (Pourpeighambar et al. 2017). Then, it was solved through a non-cooperative multi-agent learning method in which each secondary user (SU) speculated other nodes' strategies without acquisition of global information. However, these works all adopted the flat routing protocol, in which the node density and information redundancy are high. Moreover, the number of hops in the data transmission is large, which leads to large energy consumption and high cost of establishing and maintaining routing.

To improve energy efficiency and reduce network latency, hierarchical routing protocols have been studied and developed. The main idea of hierarchical routing protocols is to divide network nodes into different groups according to their geographical location and characteristic attributes,

Du, Y., Jin, H., Wang, L. and Xue, L.
A Hierarchical Routing Protocol based on Energy Consumption Weight Clustering Scheme for Cognitive Radio Networks.
DOI: 10.5220/0008868804210430
In *Proceedings of 5th International Conference on Vehicle, Mechanical and Electrical Engineering (ICVMEE 2019)*, pages 421-430
ISBN: 978-989-758-412-1

which is called clustering. The author in (Baddour et al. 2009) adopted affinity propagation to classify SUs by similarity of sensing results. Nevertheless, there were too many clusters and the cluster size was particularly small in the sensing results based clustering algorithm, which was not conducive to network management and operation. A sensing factor based clustering scheme was built in (Qi et al. 2018), which mapped the clustering problem to Constraint Maximum-Weight Edge Biclique (C-MWEB) decomposition problem. However, it only considered the maximization of cooperative sensing accuracy of the PU channels instead of the overall energy consumption of intra-cluster communication.

To improve energy efficiency and reduce network delay, in this paper we propose a hierarchical cross-layer routing protocol based on energy efficiency in multi-agent framework. Firstly, the SUs and PU channels are clustering into groups in CRN by maximizing energy consumption weight to minimize the energy consumption of intra-cluster communication. Then the multi-agent reinforcement learning algorithm is used to joint optimize the routing, channel access and power allocation of the cluster head for the reduction of transmission delay and system energy consumption. Simulation results show that the end-to-end performance of the proposed hierarchical routing scheme is significantly better than that of the flat routing protocol and the hierarchical routing protocol under the traditional clustering algorithm.

2 SYSTEM MODEL

2.1 Network Model and Frame Structure

We consider a CRN consisting of N SU nodes, in which the SUs in the set $\Lambda = \{n_1, \ldots, n_N\}$ coexist with M PUs in the set $\mathrm{P} = \{v_1, \ldots, v_M\}$, and SUs use the authorized channels for data transmission when the PU channel is idle. SUs adopt cooperative spectrum sensing to improve the sensing accuracy. The signal of PUs can only cover SUs in a certain area due to their limited transmission power. Some SU nodes located in the remote area or in the hidden terminal position cannot effectively detect the signal of PUs. Their participation in cooperative spectrum sensing may lead to a decline in sensing performance. To improve the accuracy of channel detection for PUs and reduce the number of data forwarding, we introduce a hierarchical routing protocol based on clustering. As shown in Figure 1, SUs and PU channels are divided into K clusters. In each cluster, SU nodes jointly sense the PU channels in the cluster and transmit the data to the destination node in a hierarchical form. Each cluster is denoted as $L = \{L_1, \ldots, L_K\}$, where the cluster head of the cluster L_i is represented as H_i, and the member in the cluster L_i is denoted as Π_i. In the data transmission, the member node sends data to the cluster head, and then the cluster head transmits data to the destination node via other cluster head nodes in the multi-hop manner.

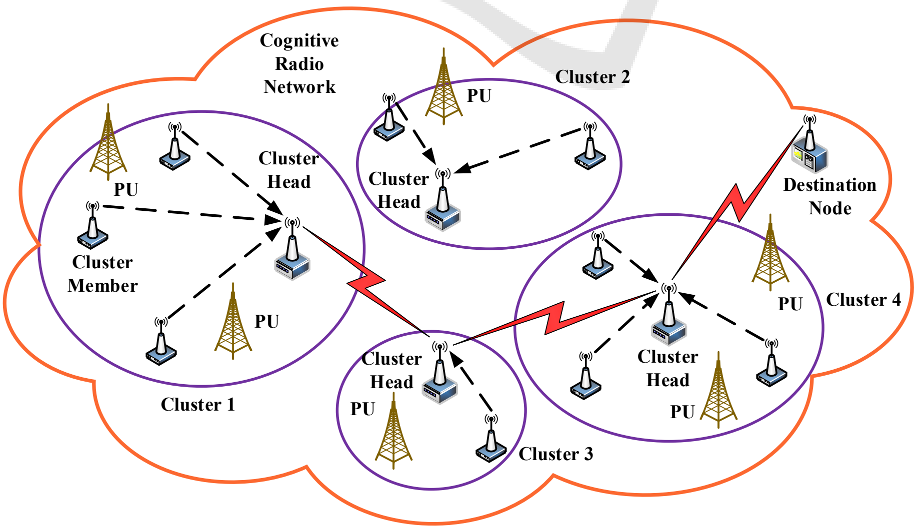


Figure 1. Network model of the hierarchical routing protocol.

As shown in Figure 2, the frame structure of hierarchical routing protocol includes clustering stage, running stage and cluster maintenance stage. In the clustering stage, SU nodes and PU channels are divided into different clusters according to certain criteria. To prevent the degradation of the original clustering performance caused by the mobility of SU nodes or the change of the PU channel characteristics, it is necessary to adjust and maintain the clustering at a certain time, which is called the cluster maintenance stage. Running phase is the most critical part of the protocol, which is divided into spectrum sensing time slot, intra-cluster data transmission time slot and inter-cluster data transmission time slot. In the spectrum sensing stage, all the nodes in the cluster execute cooperative spectrum sensing and report the sensing results to cluster head for data fusion. After that, cluster member nodes communicate with the cluster head through time division multiplexing. Subsequently, the cluster head selects the corresponding relay node and channel according to its strategy to transmit data between clusters.

In the intra-cluster data transmission stage, we assume that the transmit power of each cluster member node is a constant value P_{Π}. Since only the information of the PU channel in the cluster can be obtained, the cluster member sends the data directly to the cluster head by selecting a certain PU channel in the corresponding cluster. In the inter-cluster data transmission stage, the cluster head obtains the channel status information and the location information of the cluster head in adjacent cluster by request signaling and response signalling. Then it selects a cluster head as the relay node to forward the data to the destination node in the multi-hop manner. In this process, the transmit power of cluster head can be adaptively adjusted to reduce energy consumption and routing delay. In addition, the PU's occupation is modeled as an independently and identically alternation between two stages, i.e., ON status when the PU channel is occupied and OFF status when the PU channel is idle (Singh et al. 2017).

2.2 Energy Consumption Weight

To guarantee the system energy efficiency, the average energy consumption of data transmission within the cluster should be minimized. We assume that cluster member n_i sends data to cluster head using authorized channel j in a cluster. The channel capacity C_{ij} is defined as follows:

$$C_{ij} = B_j \log\left(1 + h_{ij} P_{\Pi} / \sigma^2\right) \tag{1}$$

Where B_j is the bandwidth of the PU channel j, h_{ij} represents the channel gain when the SU n_i using channel j, σ^2 is the Additive White Gaussian Noise (AWGN) power, and P_{Π} is the transmit power of the cluster member nodes. It is assumed that the size of the packet is S_{packet}, the data transmission time of the cluster member nodes is given by:

$$\mu_{ij} = S_{packet} / C_{ij} \tag{2}$$

Then the energy consumption in the intra-cluster data transmission is calculated as $E_{ij} = p_{\Pi}\mu_{ij}$. It is assumed that there are J PU channels in the cluster, and the idle probability of the channel j is p_{off}^{j}. We consider that the probability that a cluster member node chooses a certain channel is proportional to the idle probability of the channel. So the probability of node selecting the channel j in intra-cluster data transmission is given by:

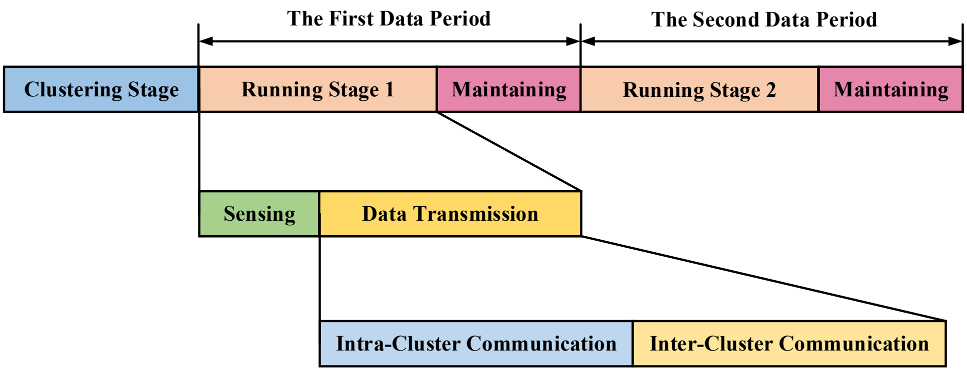


Figure 2. Frame structure of hierarchical routing protocol.

$$P_j = p_{off}^j \Big/ \sum_{n=1}^{J} p_{off}^n \tag{3}$$

Therefore, the average energy consumption of the k th cluster in the intra-cluster communication is calculated as follows:

$$E_k = \sum_i^{\Pi_k} \sum_j^{J} p_{\Pi} \mu_{ij} \cdot \left(p_{off}^j \Big/ \sum_{n=1}^{J} p_{off}^n \right) \tag{4}$$

To minimize the average energy consumption of intra-cluster communication in the whole network, it is necessary to choose a reasonable clustering method to maximize $-\sum_k E_k$. Then the concept of Energy Consumption Weight (ECW) is introduced, and the ECW of SU node n_i using PU channel j is defined as:

$$\vartheta_{ij} = -E_{ij} P_j = -p_{\Pi} \mu_{ij} \cdot \left(p_{off}^j \Big/ \sum_{n=1}^{J} p_{off}^n \right) \tag{5}$$

Thus the sum of ECW in the k th cluster is given by:

$$\begin{aligned} \Theta_k &= -E_k = \sum_i^{\Pi_k} \sum_j^{J} \vartheta_{ij} \\ &= -\sum_i^{\Pi_k} \sum_j^{J} p_{\Pi} \mu_{ij} \cdot \left(p_{off}^j \Big/ \sum_{n=1}^{J} p_{off}^n \right) \end{aligned} \tag{6}$$

3 ALGORITHMIC DESIGN

3.1 Energy Consumption Weight Based Clustering Algorithm

We introduce the concept of bipartite graph in graph theory to cluster SU nodes and PU channels reasonably. As shown in Figure 3, the graph G is defined as a bipartite graph if the vertices of undirected graph $G(X \cup Y, \delta)$ can be divided into two independent sets, and the two vertices x and y connected by the edge in δ belong to the set X and Y, respectively. Vertex set X corresponds to the set of SU nodes in CRN, and the set Y represents the PU channels. If a SU node x is within the coverage of the PU y and can sense the state of the corresponding PU channel, then there is an edge (x, y) in the bipartite graph where the weight of the edge is the ECW ϑ_{ij}.

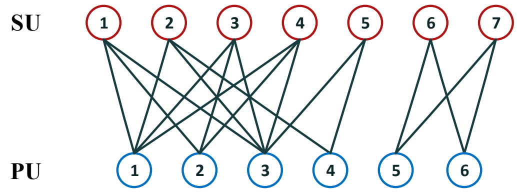


Figure 3. Bipartite graph model.

For a bipartite graph $Q(U, W)$, if there is an edge connection between any two vertices $u \in U$ and $w \in W$, then Q is called a complete bipartite graph. For CRN, the problem of clustering SU nodes and PU channels can be formulated as the problem of decomposing a bipartite graph into a complete bipartite graph. Figure 4 shows three cases in which the bipartite graph in Figure 3 is decomposed into a complete bipartite graph. The complete bipartite graph shown in Figure 4(a) represents the first cluster partitioned from the CRN. Its cluster member set is $U_1 = \{1,3,4\}$, and the sensed channel set is $W_1 = \{1,2,3\}$. All of the SU nodes in U_1 sense the channels in W_1 cooperatively and communicate with cluster head through these channels. Similarly, the complete bipartite graph shown in Figure 4(b) denotes the second cluster. Its cluster member set is $U_2 = \{6,7\}$, and the sensed channel set is $W_2 = \{5,6\}$. The complete bipartite graph shown in Figure 4(c) denotes the third cluster. Its cluster member set is $U_3 = \{2,5\}$, and the sensed channel set is $W_3 = \{4\}$.

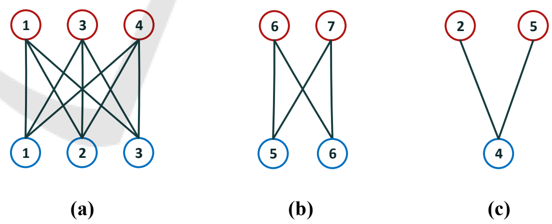


Figure 4. Complete bipartite graph decomposed from the bipartite graph in Figure 3.

To minimize the average energy consumption of intra-cluster communication after clustering, it is necessary to maximize the sum of system ECW. Therefore, we have to find a bipartite graph decomposition method to maximize the sum of edge weights of all complete bipartite graphs after decomposition. The matrix A_s and B_p represent the clustering of SU nodes and PU channels, respectively. The elements a_i^k denotes the relationship between SU node n_i and the k th

cluster, and the elements b_j^k represents the relationship between the PU v_i and the k *th* cluster:

$$a_i^k = \begin{cases} 1, & n_i \in U_k \\ 0, & else \end{cases} \tag{7}$$

$$b_j^k = \begin{cases} 1, & v_i \in W_k \\ 0, & else \end{cases} \tag{8}$$

Consequently, the clustering problem can be mapped to the problem of bipartite graph decomposition to maximize the weight of edge in graph theory. The mathematical description is as follows:

$$\begin{aligned} &\max_{A_s, B_p} \sum_k \Theta_k(U_k(A_s), W_k(B_p)) \\ &\text{s.t.} \sum_{k=1}^{K} a_i^k = 1 \quad \forall i \\ &\quad\;\; \sum_{k=1}^{K} b_j^k = 1 \quad \forall j \end{aligned} \tag{9}$$

Two constraints in Equation (9) indicate that each SU node or PU channel can only be allocated to one cluster and cannot appear in different clusters. In addition, it is assumed that the number of SU nodes in each cluster is not less than m to guarantee cooperative spectrum sensing accuracy.

Since solving the optimization objective (9) is a NP-complete problem (Zhang et al. 2014), a heuristic greedy algorithm is designed to find the next optimal solution. The algorithm can effectively separate the complete bipartite graph from the bipartite graph under the above constraints. In the initial stage, it is assumed that $\Lambda_1 = \Lambda$ and $\mathrm{P}_1 = \mathrm{P}$. For the k *th* cluster, $\Lambda_k = \Lambda_{k-1}/U_{k-1}$ and $\mathrm{P}_k = \mathrm{P}_{k-1}/W_{k-1}$. The steps to obtain a complete bipartite graph $Q(U_k, W_k)$ from a bipartite graph $G(\Lambda_k \cup \mathrm{P}_k, \delta_k)$ are as follows: W_k is set to an empty set and U_k is set to the SU node set. In the l *th* iteration, we first find the PU channel $v_l \in \mathrm{P}_k/W_k$ with the highest edge weights, whose edge weights are denoted as $\deg(v_l)$. Then we add v_l to W_k and remove the SU node that cannot sense the channel v_l from U_k (The set of SU nodes that can sense v_l is denoted as ψ_{v_l}). Subsequently, we calculate the sum of all the edge weights of the bipartite graphs composed of U_k and W_k, which is denoted as Γ_k . The above steps are repeated until the number of remaining channels in P_k is 0 or $|U_k| < m_k$. The details of greedy clustering algorithm based on ECW are shown in algorithm 1.

Algorithm 1 Greedy clustering algorithm based on energy consumption weight

1: **Initialize:**
2: Set $G(\Lambda_k \cup \mathrm{P}_k, \delta_k)$, $U_k = \Lambda$, $W_k = \phi$ and $l \leftarrow 1$;
3: **While** $|U_k| \geq m_k$ and $|\mathrm{P}_k| > 0$ **Do**
4: $v_l = \arg\max_{v \in \mathrm{P}_k/W_k} \deg(v)$;
5: **If** $\deg(v) < m_k$ **then**
6: break;
7: **Else**
8: $\mathrm{P}_k \leftarrow \mathrm{P}_k - v_l$; $U_k \leftarrow U_k \cap \psi_{v_l}$;
9: $\psi_{v_l} = \{SU \in \Lambda_k \,|\, SUs \text{ that can senses } v_l\}$;
10: $W_k \leftarrow W_k \cup v_l$; $\Gamma_k[l] = \Theta_k(U_k, W_k)$;
11: **End If**
12: $l = l + 1$
13: **End While**
14: **Output:** $Q_k^*(U_k, W_k)$

3.2 Inter-Cluster Communication Protocol

After clustering, we select cluster heads using ID (LID) scheme which is commonly used in ad hoc networks. All cluster heads form the upper network structure, and transmit data derived from the source node to the destination node with the multi-hop manner. In this paper, the inter-cluster cross-layer routing design problem is modeled as a quasi-cooperative stochastic game, and the concept of responsibility rating proposed in (Du et al. 2018) is applied. The players of the stochastic game are all cluster heads. In the game, each player chooses an action and then obtains a reward based on the present state and action. Subsequently, the game will enter the next stage and its state is determined by the previous state and the action of each players. In the stochastic game, the state distribution, player's action and reward at each stage are defined as follows:

(1) State distribution:

The state of cluster head H_i at timeslot t is $s_i^t = \{\rho_i, \varphi_i^t, f_i^t\}$, where φ_i^t is the responsibility rating of the cluster head H_i, $f_i^t \in \mathbf{C_i}$ is the PU channel accessed by the cluster head H_i at timeslot t, and

$\rho_i \in \{0,1\}$ is the Signal-to-Interference plus Noise Ratio (SINR) indicator that indicates whether the SINR γ_i of cluster head H_i is above or below the threshold γ_{th} :

$$\rho_i = \begin{cases} 1, & \text{if } \gamma_i \geq \gamma_{th} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Where $\gamma_i = g_{ijc} p_{\varphi_i^t} \big/ \left(\sigma^2 + \phi_{ijc}^{PU}\right)$, $p_{\varphi_i^t}$ is the transmitting power of cluster head H_i , g_{ijc} represents the channel gain between node n_i and n_j , ϕ_{ijc}^{PU} denotes the PU-to-SU interference at H_i , and σ^2 is the AWGN power. In addition, a learning episode of cluster head H_i terminates when $\rho_i = 0$, i.e., $s_i^t = \{0, \varphi_i^t, f_i^t\}$ is the terminal state in the Markov chain.

(2) Player's action:

Player's action consists of routing selection, channel access and power control of cluster head nodes. The action of cluster head H_i at timeslot t is $a_i^t = \{H_j, c_i, p_{\varphi_i}\}$, where H_j represents the relay cluster head selected form the neighbouring node of cluster head H_i , $c_i \in \mathbf{C_i}$ represents the PU channel of cluster head H_i , and p_{φ_i} is the transmit power of cluster head H_i corresponding to the responsibility rating φ_i^t .

(3) Instantaneous reward:

The single-hop transmission latency $U_{TL,i}$ of the cluster head H_i is defined as follows:

$$U_{TL,i} = S_{packet} \big/ \left[B_j \cdot \log_2\left(1+\gamma_i\right)\right] \tag{11}$$

Where S_{packet} represents the data packet size, B_j is the bandwidth of PU channel j , and γ_i is the SINR of cluster head H_i . The power consumption ratio $U_{PCR,i}$ of the cluster head H_i is given by:

$$U_{PCR,i} = p_{\varphi_i} \big/ \left[B_j \cdot \log_2\left(1+\gamma_i\right)\right] \tag{12}$$

Where p_{φ_i} is the transmit power of cluster head H_i corresponding to the responsibility rating φ_i^t . The instantaneous reward when cluster head H_i executes action a_i in s_i and other cluster heads perform actions $\mathbf{a_{\text{-}i}}$ is defined as:

$$R_i^t(s_i, a_i, \mathbf{a_{\text{-}i}}) = \text{-}\log_2\left(\alpha \cdot U_{TL,i} + \beta \cdot U_{PCR,i}\right) \tag{13}$$

Where $\mathbf{a_{\text{-}i}} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_K) \in \mathbf{A_{\text{-}i}} = \prod_{j \in H \setminus H_i} \mathbf{A_j}$ is other cluster heads' action vector; α and β are parameters to adjust the weighting of the transmission delay and energy efficiency.

Each cluster head only needs local information instead of sharing information with all cluster heads in the network. Therefore, the cross-layer design problem of inter-cluster communication can be modeled as a non-cooperative stochastic game:

$$\begin{aligned} &\max_{a_i \in \mathbf{A_i}} R_i^t(s_i, a_i, \mathbf{a_{\text{-}i}}) \\ s.t. \quad & U_{TD,i} \leq \kappa_{th} \end{aligned} \tag{14}$$

Where κ_{th} is the maximum transmission latency of the cluster head.

We apply the Equal Reward Timeslots based Conjectural Multi-Agent Q-Learning (ERT-CMAQL) proposed in (Du et al. 2019) to solve the Nash equilibrium of the stochastic game and optimize the routing and resource allocation in the inter-cluster communication.

The update framework of multi-agent Q-learning is given by:

$$Q_i^{t+1}(s_i, a_i) = (1-\alpha) Q_i^t(s_i, a_i) + \alpha \begin{bmatrix} E\left[R_i(s_i, \psi_i, \mathbf{\Psi_{\text{-}i}})\right] \\ +\beta \max_{b_i \in \mathbf{A_i}} Q_i^t(s_i', b_i) \end{bmatrix} \tag{15}$$

Where $\alpha \in [0, 1)$ is the learning rate, and β is the discount factor determining the agent's horizon, $E\left[R_i(s_i, \psi_i, \mathbf{\Psi_{\text{-}i}})\right]$ is the expected reward for cluster head H_i at timeslot t considering other $N-1$ competing cluster heads, that is

$$E\left[R_i(s_i, \psi_i, \mathbf{\Psi_{\text{-}i}})\right] = \sum_{(a_i, \mathbf{a_{\text{-}i}}) \in \mathbf{A}} \left[R_i(s_i, a_i, \mathbf{a_{\text{-}i}}) \prod_{j \in H \setminus H_i} \psi_j(s_j, a_j)\right] \tag{16}$$

Where $\psi_j(s_j, a_j)$ is the strategy of cluster head H_j . Thus the mixed-strategies for other cluster heads is defined as:

$$\chi_i^t(s_i, \mathbf{a_{\text{-}i}}) = \prod\nolimits_{j \in H \setminus H_i} \psi_j^t(s_j, a_j) \tag{17}$$

Which represents the probability that other cluster heads execute strategy $\mathbf{\Psi_{\text{-}i}}$ at timeslot t. Furthermore, the probability that the agent chooses

a_i at state s_i while other competing cluster heads performing strategy $\boldsymbol{\psi}_{\text{-i}}$ is given by:

$$\phi_i = \psi_i^t(s_i, a_i) \cdot \chi_i^t(s_i, \mathbf{a}_{\text{-i}}) \tag{18}$$

That is, the probability that cluster head H_i obtains expected reward $R_i(s_i, a_i, \mathbf{a}_{\text{-i}})$ is ϕ_i. Let n denotes the number of time slots between any two continuous slots which cluster head H_i gets the same return $R_i(s_i, a_i, \mathbf{a}_{\text{-i}})$, and n has the independent and identical distribution with ϕ_i. It is assumed that the average value of n is denoted as $\bar{n}$, and then we have the approximate equation $\phi_i \approx 1/(1+\bar{n})$. Since every cluster head has its own strategy $\psi_i^t(s_i, a_i)$, the agent is able to estimate $\chi_i^t(s_i, \mathbf{a}_{\text{-i}})$ as follows:

$$\chi_i^t(s_i, \mathbf{a}_{\text{-i}}) = 1\Big/\left[(1+\bar{n}) \cdot \psi_i^t(s_i, a_i)\right] \tag{19}$$

Since n is a stationary stochastic process in time dimension so its mean value $\bar{n}$ is a constant. Specifically, the quotient of the conjecture belief at time slot t - 1and t can be calculated as:

$$\begin{aligned}\frac{\chi_i^t(s_i, \mathbf{a}_{\text{-i}})}{\chi_i^{t-1}(s_i, \mathbf{a}_{\text{-i}})} &= \frac{1}{(1+\bar{n}) \cdot \psi_i^t(s_i, a_i)}\Bigg/\left[\frac{1}{(1+\bar{n}) \cdot \psi_i^{t-1}(s_i, a_i)}\right] \\ &= \frac{\psi_i^{t-1}(s_i, a_i)}{\psi_i^t(s_i, a_i)}\end{aligned} \tag{20}$$

Thus $\chi_i^t(s_i, \mathbf{a}_{\text{-i}})$ is updated as follows:

$$\chi_i^t(s_i, \mathbf{a}_{\text{-i}}) = \chi_i^{t-1}(s_i, \mathbf{a}_{\text{-i}}) \cdot \frac{\psi_i^{t-1}(s_i, a_i)}{\psi_i^t(s_i, a_i)} \tag{21}$$

Consequently, the multi-agent Q-learning updating rule in (15) can be modified as:

$$\begin{aligned}&Q_i^{t+1}(s_i, a_i) = (1-\alpha) Q_i^t(s_i, a_i) \\ &+\alpha\left[\sum_{(a_i, \mathbf{a}_{\text{-i}}) \in \mathbf{A}} R_i(s_i, a_i, \mathbf{a}_{\text{-i}}) \chi_i^t(s_i, \mathbf{a}_{\text{-i}}) + \beta \max_{b_i \in \mathbf{A}_i} Q_i^t(s_i', b_i)\right]\end{aligned} \tag{22}$$

The details of ERT-CMAQL based cross-layer routing protocol for inter-cluster communication are shown in algorithm 2.

Algorithm 2 ERT-CMAQL based cross-layer routing protocol for inter-cluster communication

Initialize:
Set $t = 0$ and memory size N.
For each cluster head H_i **Do**
  For each $s_i \in S_i$, $a_i \in A_i$ **Do**
    Initialize $\psi_i^t(s_i, a_i)$, $\chi_i^t(s_i, \mathbf{a}_{\text{-i}})$, $Q_i^t(s_i, a_i)$.
  End For
End For
Learning:
For each cluster head H_i **Do**
  For $eposide = 1, M$ **Do**
    Initialize state s_i^1.
    Repeat
      Select action a_i^t according to $\psi_i^t(s_i, a_i)$.
      Execute a_i^t, and obtain ρ_i.
      Observe $R_i^t(s_i, a_i, \mathbf{a}_{\text{-i}})$ and s_i^{t+1}.
      Update $Q_i^{t+1}(s_i, a_i)$ based on $\chi_i^t(s_i, \mathbf{a}_{\text{-i}})$ according to (22).
      Update the strategy $\psi_i^{t+1}(s_i, a_i)$ according to Boltzmann distribution.
      Update $\chi_i^{t+1}(s_i, \mathbf{a}_{\text{-i}})$ according to (21).
      $s_i = s_i^{t+1}$
      $t = t + 1$
    Until s_i is the terminal state
  End For
End For

4 SIMULATION RESULTS

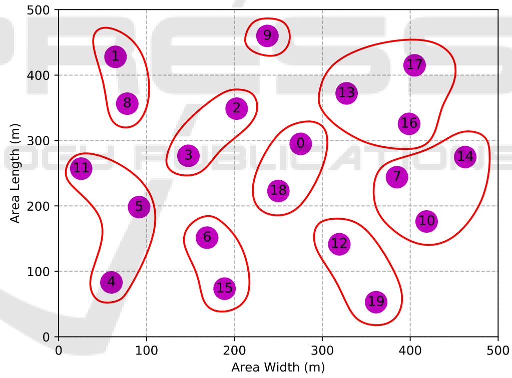
In this section, the performance of the hierarchical routing protocol based on ECW clustering algorithm is evaluated using Python 3.5.1 and its packages Networkx 2.3 and Numpy 1.15.3. The results of the proposed scheme are compared with (1) Cooperative Multi-Agent Q-Learning (CMAQL) under flat routing protocol (Du et al. 2019); (2) ERT-CMAQL based on C-MWEB clustering algorithm (Qi et al. 2018) and (3) Q-routing (Al-Rawi et al. 2014) based on ECW clustering algorithm.

In the simulation process, the bandwidth of PU channel j is set to $B_j \sim [1, 2]$ MHz . It is supposed that the AWGN power $\sigma^2 = 10^{-7}$ mW , the packet size $S_{packet} = 2\times10^5$ bit , and the PU-to-SU interference $\phi_{ijc}^{PU} \sim [10^{-7}, 10^{-6}]$ mW . The link gain of both intra-cluster communication and inter-cluster communication is given by:

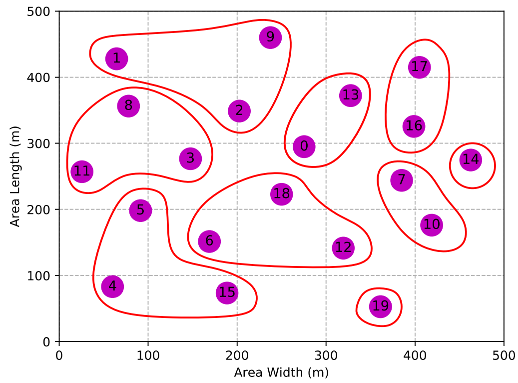
$$h = LF(d/d_0)^{-q} \quad , \text{for } d > d_0 \qquad (23)$$

Where L is a constant set to be 10^{-6} , and the shadowing factor F is subject to a lognormal distribution with a mean of $0\,dB$ and variance of $6\,dB$, d is the actual distance between the transmitter and receiver, d_0 is the reference distance, and q is the path loss exponent. In our simulation process, we set $d_0 = 1$ and $q = 4$. The expected mean and deviation of average PU departure rate $\theta_d(\mu, \vartheta)$ are set as $\mu = 0.1$ and $\vartheta = 0.05$. A networking scenario comprising 20 SUs and 10 PUs uniformly deployed in a 500 × 500 m area is considered. The transmit power of cluster member node is $P_{\Pi} = 200$mW , and the transmit power set for cluster head contains ten levels: $\{550, 600, \ldots, 1000 \text{ mW}\}$. We use ECW based clustering scheme and C-MWEB clustering algorithm to cluster the network. The C-MWEB clustering algorithm only considers the maximization of cooperative sensing accuracy of PU channels, but it ignores the overall energy consumption of intra-cluster communication. The network topology and clustering results are shown in Figure 5.

Figure 6 shows the packet transmission delay varying with the number of routes in different schemes. It can be seen that with increasing number of routes, the packet transmission delay of all schemes decreases gradually. This is because agents gradually learn the optimal strategies under each scheme by interacting with the environment. As a single agent scheme, the packet transmission delay of Q-routing algorithm should be much higher than that of the multi-agent strategy CMAQL. However, we can see that the packet transmission delay of Q-routing algorithm based on ECW clustering is lower than that of CMAQL algorithm under flat routing protocol. The reason for this is that the hierarchical routing protocol reduces the number of data forwarding. Furthermore, the transmit power of cluster head is larger than that of flat routing protocol so that the data transmission time of each hop is reduced. Thus the total transmission delay of Q-routing algorithm under hierarchical routing protocol is much lower. In addition, when ERT-CMAQL is adopted as the inter-cluster communication scheme, the packet transmission delay of the hierarchical routing protocol based on ECW is lower than that of the C-MWEB clustering protocol. This is because the average channel capacity of each SU in the cluster is larger so that the transmission delay in the intra-cluster communication is lower than that in the C-MWEB clustering scheme, which leads to lower total transmission delay. Moreover, we find that ERT-CMAQL based on ECW clustering, ERT-CMAQL based on C-MWEB clustering and CMAQL under flat routing protocol have almost the same convergence speed, which is faster than Q-routing based on ECW clustering. This is due to the fact that each SU node in the multi-agent learning scheme is equipped with an agent, and each agent works in parallel so that the convergence rate is not affected by the size of the network. In the single agent framework, one agent works independently and its computation load is heavy so that the convergence speed is slower than that of multi-agent system.



(a) ECW based clustering scheme.



(b) C-MWEB clustering algorithm.

Figure 5. Network topology and comparison of clustering results.

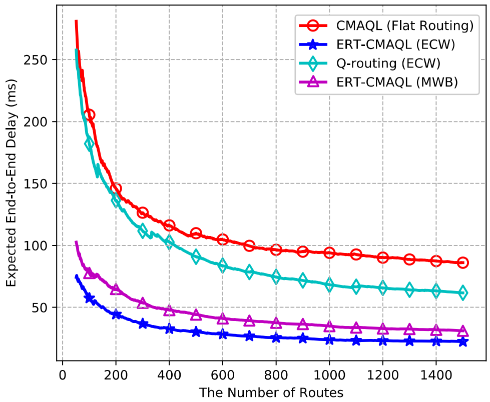


Figure 6. Data packet delay vs. the number of routes.

Figure 7 illustrates the power consumption ratio changing with the number of routes. It can be seen that the power consumption ratio of Q-routing based on ECW clustering is lower than that of CMAQL under flat routing protocol. This is mainly because the hierarchical routing protocol reduces the times of data forwarding. Although the transmit power of cluster heads is higher than that of nodes in flat routing, the advantages of less data forwarding times will not be offset because the packet size is sufficiently large. In addition, when ERT-CMAQL is adopted as the inter-cluster communication scheme, the energy consumption of the hierarchical routing protocol based on ECW is lower than that of the C-MWEB clustering protocol. This is because the C-MWEB clustering algorithm only considers the maximization of cooperative sensing accuracy of PU channels, but it ignores the overall energy consumption of intra-cluster communication. Therefore, the average energy consumption in the intra-cluster communication is larger when using C-MWEB clustering algorithm so that the system power consumption ratio is higher.

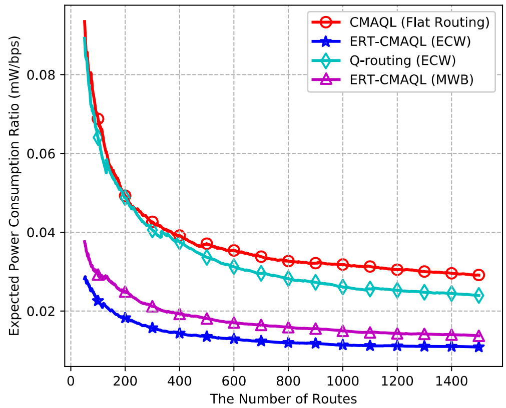


Figure 7. Power consumption ratio vs. the number of routes.

As shown in Figure 8, the Packet Loss Rate (PLR) of all algorithms decreases gradually with the increase of the number of routes, and the PLR of Q-routing based on ECW clustering is lower than that of CMAQL under flat routing protocol. In addition, when ERT-CMAQL is adopted as the inter-cluster communication scheme, the PLR of hierarchical routing protocol based on ECW is lower than that of the C-MWEB clustering protocol. This is because hierarchical routing protocol based on ECW has lower packet transmission delay in Figure 6. Then the number of packets whose total transmission latency exceeds the delay tolerance is smaller so that the number of packets which are transmitted successfully exceeds the protocol based on C-MWEB clustering. This shows that ERT-CMAQL based on ECW clustering has higher network stability than other algorithms.

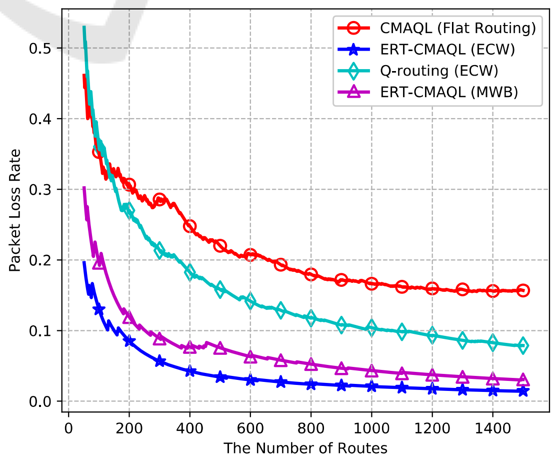


Figure 8. Packet loss rate vs. the number of routes.

5 CONCLUSIONS

In this paper, we developed a hierarchical routing protocol based on energy consumption weight clustering scheme. Firstly, SU nodes and PU channels in CRN are clustered by maximizing energy consumption weights for the minimization of the energy consumption in intra-cluster communication. Then the strategy conjecture based multi-agent Q-learning scheme is used to joint optimize the routing, channel access and power allocation of the cluster head for the reduction of transmission delay and system energy consumption. Simulation results show that the end-to-end performance of the proposed hierarchical routing scheme is significantly better than that of the flat routing protocol and the hierarchical routing protocol under the traditional clustering algorithm.

REFERENCES

Al-Rawi, H. A., Yau, K. L. A., Mohamad, H., Ramli, N., & Hashim, W., 2014. A reinforcement learning-based routing scheme for cognitive radio ad hoc networks. *In 2014 7th IFIP wireless and mobile networking conference (WMNC)* (pp. 1-8). IEEE.

Baddour, K. E., Ureten, O., & Willink, T. J., 2009. Efficient clustering of cognitive radio networks using affinity propagation. *In 2009 Proceedings of 18th International Conference on Computer Communications and Networks (pp. 1-6)*. IEEE.

Cao, Y., Duan, D., Cheng, X., Yang, L., & Wei, J., 2014. QoS-oriented wireless routing for smart meter data collection: Stochastic learning on graph. *IEEE Transactions on Wireless Communications*, 13(8), 4470-4482.

Cesana, M., Cuomo, F., & Ekici, E. (2011). Routing in cognitive radio networks: Challenges and solutions. *Ad Hoc Networks*, 9(3), 228-248.

Chen, H., Zhou, M., Xie, L., Wang, K., & Li, J., 2016. Joint spectrum sensing and resource allocation scheme in cognitive radio networks with spectrum sensing data falsification attack. *IEEE Transactions on Vehicular Technology*, 65(11), 9181-9191.

Du, Y., Chen, C., Ma, P., & Xue, L., 2019. A Cross-Layer Routing Protocol Based on Quasi-Cooperative Multi-Agent Learning for Multi-Hop Cognitive Radio Networks. *Sensors*, 19(1), 151.

Du, Y., Zhang, F., & Xue, L., 2018. A kind of joint routing and resource allocation scheme based on prioritized memories-deep Q network for cognitive radio ad hoc networks. *Sensors*, 18(7), 2119.

Pourpeighambar, B., Dehghan, M., & Sabaei, M., 2017. Non-cooperative reinforcement learning based routing in cognitive radio networks. Computer communications, 106, 11-23.

Qi, Q., Wang, K., & Du Y., 2018. A Clustering Scheme based on Spectrum Sensing in Cognitive Radio Ad hoc Networks. *Journal of Data Acquisition and Processing* (1), 41-50.

Singh, K., & Moh, S., 2017. An Energy-Efficient and Robust Multipath Routing Protocol for Cognitive Radio Ad Hoc Networks. *Sensors*, 17(9), 2027.

Zhang, W., Yang, Y., & Yeo, C. K., 2014. Cluster-based cooperative spectrum sensing assignment strategy for heterogeneous cognitive radio network. *IEEE Transactions on Vehicular Technology*, 64(6), 2637-2647.