Journal of Network Intelligence ©2023 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **8**, Number **4**, November **2023**

A Simultaneous Wireless Information and Power Transfer-based Multihop Uneven Clustering Routing Protocol for Cognitive Radio Sensor Networks

Ji-Hong Wang*

School of Electrical Engineering
Northeast Electric Power University
169 Changchun Road, Jilin, 132012, China
wangjihong07@126.com

Li-Dong Zhang

School of Electrical Engineering
Northeast Electric Power University
169 Changchun Road, Jilin, 132012, China
zhanglidong1010@126.com

*Corresponding author: Ji-Hong Wang

Received November 8, 2022, revised January 4, 2023, accepted March 13, 2023.

Abstract. *Multihop clustering routing protocols are potential solutions to achieve effective data delivery in large-scale cognitive radio sensor networks (CRSNs). Current clustering routing protocols for CRSNs may cause unbalanced energy distribution among nodes and even energy holes. In order to solve this problem, a simultaneous wireless information and power transfer (SWIPT)-based multihop uneven clustering routing protocol (S-MUCRP) is proposed for CRSNs in this paper. The energy consumption of control overhead and data transmission is comprehensively taken into consideration, and cluster radii are theoretically derived with the purpose of minimizing the energy consumption of the outmost layer and balancing the net energy consumption among cluster heads (CHs) in different layers. Energy level function-based CHs and relay selection criteria are defined to help determine high-quality CHs and relays and improve energy sustainability and network connectivity. Intra-cluster SWIPT is introduced to further balance the residual energy among nodes in the same cluster. Simulation results show that compared with the existing clustering routing protocols for CRSNs, S-MUCRP can guarantee effective network surveillance during long network lifetime.*
Keywords: Cognitive radio sensor networks, Uneven clustering, Multihop routing, Simultaneous wireless information and power transfer

---

1. **Introduction.** Wireless sensor networks (WSNs) are self-organizing networks composed of a large number of sensor nodes which collaborate with each other, and they are crucial components of Internet of Things [1, 2]. These sensor nodes perceive the environment and transmit the perceived information towards the sink [3, 4]. However, with the development of various wireless communication techniques such as Wi-Fi, Bluetooth and Zigbee, WSNs are faced with spectrum scarcity problem which seriously restricts the network performance. As an effective solution to solve the spectrum shortage problem, cognitive radio (CR) is introduced into WSNs and they are intelligently combined to form cognitive radio sensor networks (CRSNs) [5, 6]. Inherited from WSNs, CRSNs nodes are

generally powered by limited-capacity battery [7, 8], and they also need to consume extra energy to perform CR functions which will further accelerate their energy exhaustion [9]. Therefore, CRSNs are faced with severe energy constraint problem.

Clustering protocols can eliminate data redundancy and shorten the communication distance, correspondingly, they have great potentials to reduce node energy consumption and prolong the network lifetime [10]. In recent years, multihop clustering routing protocol design for CRSNs has attracted the attention from both academia and industry, and it has become a hot research topic. However, the existing clustering routing protocols for CRSNs suffer the following limitations: (1) Cluster heads (CHs) close to the sink are required to perform heavy data forwarding tasks which will quickly drain their battery. Therefore, uniform clustering routing protocols may result in energy holes or even network partition [11]. (2) Uneven clustering routing protocols form clusters with different size to balance the energy distribution among nodes as much as possible [12]. Current uneven clustering protocols generally adopt constant coefficients and linear relationship with the Euclidean distance to the sink to quantify the cluster radius. Actually, the above coefficients and cluster radius should be determined according to specific network configurations. (3) Focusing only on the energy consumption of data communication, the impact of control overhead is left out of consideration, which may result in inaccurate analysis results.

In order to conquer the above limitations and make full use of uneven clustering to balance the residual energy among nodes, a simultaneous wireless information and power transfer (SWIPT)-based multihop uneven clustering routing protocol (S-MUCRP) is proposed in this paper. Cluster radii are theoretically derived based on specific network configurations to manage the range of control information exchange, which can effectively reduce the energy consumption. High-quality CHs and relays are chosen by applying energy level function-based criteria, which can help enhance the energy sustainability and network connectivity. Intra-cluster SWIPT is introduced to further balance the residual energy among nodes and prolong the network lifetime.

2. Related works.

Based on cluster size, current clustering protocols for CRSNs can roughly be divided into uniform and uneven clustering protocols. In addition, the academy and industry introduced SWIPT [13] into clustering protocols to further balance the residual energy among nodes. Therefore, the related works will be reviewed from 3 aspects, that is, uniform clustering protocols for CRSNs, uneven clustering protocols for CRSNs and SWIPT-based clustering protocol design. The characteristics analysis of these protocols is shown in Table 1.

2.1. Uniform clustering protocols for CRSNs.

Uniform clustering protocols for CRSNs can be further categorized into centralized, distributed and hybrid clustering protocols. Fuzzy C-means [14], IMOCRP [15] and CogLEACH-C [16] are representatives of centralized clustering protocols. To be specific, Fuzzy C-means divides nodes into clusters with the purpose of minimizing the summation of the squared distance between cluster members (CMs) and the cluster center. CHs are selected based on the relative location to the cluster center, signal to noise ratio and node residual energy. IMOCRP is a time-triggered clustering routing protocol based on ions motion optimization algorithm, and it automatically determines the optimal number of clusters and CHs. NonCHs nodes join clusters according to the Euclidean distance and common available channels. However, the network scalability is restricted, and it is unsuitable for large-scale CRSNs. CogLEACH-C selects CHs according to idle channels and node residual energy. Typical distributed clustering protocols include CogLEACH [17], DSAC [18], SCEEM [19],

Table 1. Characteristics analysis and comparison of related works

Protocols	Target network	Types	Whether the impact of control overhead is considered	Whether SWIPT is applied
Fuzzy C-means	CRSNs	Centralized uniform clustering	×	×
IMOCRP	CRSNs	Centralized uniform clustering	✓	×
CogLEACH-C	CRSNs	Centralized uniform clustering	×	×
CogLEACH	CRSNs	Distributed uniform clustering	×	×
DSAC	CRSNs	Distributed uniform clustering	×	×
SCEEM	CRSNs	Distributed uniform clustering	×	×
NSAC	CRSNs	Distributed uniform clustering	×	×
EACRP	CRSNs	Distributed uniform clustering	×	×
WCM	CRSNs	Hybrid uniform clustering	×	×
LEAUCH	CRSNs	Uneven clustering	×	×
OACUCAPTEEN	CRSNs	Uneven clustering	×	×
ESAUC	CRSNs	Uneven clustering	×	×
R-bUCRP	CRSNs	Uneven clustering	×	×
SWIPT-based ENO	WSNs	Centralized uniform clustering	×	✓
CREST	WSNs	Uneven clustering	×	✓
S-MUCRP	CRSNs	Uneven clustering	✓	✓

NSAC [20] and EACRP [21]. CogLEACH adopts the number of idle channels as a main metric for CHs selection. Each CH broadcasts temporary and final CHs notification message, while nonCHs nodes unicast temporary request and final confirmation message to the CH which possesses the lowest communication cost. In DSAC, each node acts as a separate CH at the beginning, and then they are merged with each other according to common available channels and inter-cluster distance until the theoretically optimal number of clusters is achieved. Although DSAC enables strong network scalability and stability, continuous cluster merging accelerates node energy exhaustion, which may shorten the network lifetime. Spectrum-aware energy-efficient multimedia clustering routing protocol SCEEM theoretically derives the optimal number of clusters by minimizing the source loss, and it determines CHs based on node residual energy, channel idle ratio and average channel available time. In NSAC, all CRSNs nodes calculate their CHs weight based on residual energy and channel quality, and they constantly update and broadcast the weight information. The node which possesses the highest weight in the locality becomes a final CH, and neighboring nodes join its cluster. The above process is repeated until all nodes are clustered. EACRP chooses nodes with more residual energy, more common available channels with neighbors and closer to the sink as gateway nodes. In addition, CHs rotation is leveraged to change CHs round by round. Hybrid clustering protocol WCM obtains the optimal number of clusters by solving the corresponding optimization problem, and it selects CHs according to time-spatial correlation, confidence level and node residual energy [22]. After clustering, CHs substitute their CMs to perform spectrum sensing, which can significantly reduce energy consumption and increase data transmission opportunity.

In large-scale CRSNs which adopt uniform clustering protocols, all data needs to be eventually delivered to the sink. Therefore, CHs close to the sink need to relay more data and their energy exhaustion is faster, which may result in energy holes or even network partition.

2.2. **Uneven clustering protocols for CRSNs.** Uneven clustering protocols aim at solving the energy hole problem by adjusting the cluster size and balancing the energy consumption among CHs. LEAUCH calculates CHs weight and selects CHs based on number of idle channels [23]. Cluster radius is calculated according to Equation (1), and nodes with the maximum amount of energy become final CHs. CMs join corresponding clusters according to the Euclidean distance and common available channels. Optimized ant colony-based adaptive clustering protocol OACUCAPTEEN adopts the same manner as LEAUCH to calculate cluster radius, and it selects candidate CHs according to the expected number of CHs, number of licensed channels, number of idle available channels and node residual energy [24]. Candidate CHs with the maximum amount of energy become final CHs. Energy and spectrum-aware uneven clustering protocol ESAUC further takes node residual energy, number of neighbors and idle channel probability into account to calculate cluster radius on the basis of LEAUCH [25]. Reputation-based uneven clustering routing protocol R-bUCRP introduces reputation mechanism to assist in inter-node cooperation to recognize selfish nodes [26]. It calculates cluster radius according to Equation (1), and the optimal CHs are selected within the cluster radius based on residual energy and reputation of candidate CHs.

$$R_c = \left(1 - c\frac{d_{max} - d_{tosink}(i)}{d_{max} - d_{min}}\right) R_c^0 \tag{1}$$

where c is an uneven clustering coefficient; d_{max} and d_{min} are the maximum and minimum distance from all CRSNs nodes to the sink, respectively. $d_{tosink}(i)$ is the Euclidean distance from node i to the sink; R_c^0 is the largest cluster radius.

The above uneven clustering routing protocols mainly calculate cluster radius according to the Euclidean distance to the sink, and the uneven clustering coefficient c is fixed. Actually, it should be optimized according to specific network configurations.

2.3. **SWIPT-based clustering protocol design.** In SWIPT-based ENO scheme, CMs transmit data and residual energy simultaneously to their CH in a cooperative manner [27]. To guarantee neutral operation and improve the achievable data transmission rate as much as possible, the optimal SWIPT ratio is optimized with the objective of maximizing the minimum achievable data transmission rate. Basic k-means algorithm is improved to facilitate reasonable clustering and CHs selection. An SWIPT-based distributed cross-layer design scheme CREST is proposed for energy-neutral WSNs [28]. Even and uneven annular track division methods are proposed, and the track width is determined by the Euclidean distance to the sink and node density. CREST determines cluster radius with the purpose of balancing the load among all CHs in the network.

As stated above, SWIPT-based clustering protocol design is still in its infancy, and current research results are only suitable for WSNs, as they leave the impact of dynamic spectrum availability out of consideration. Current research results mainly focus on neutral operation, i.e., they assume that the harvested energy is always more than the corresponding energy consumption. In this case, all the excessive energy should be delivered to CHs. Actually, the amount of harvested energy is usually less than node energy consumption, and whether nodes will leverage SWIPT to transmit energy together with data and how much energy they will transmit should be determined according to their energy potential and the request of CHs. In addition, current clustering protocols only consider about data communication and leave the energy consumption of control overhead out of consideration, which will result in unrealistic analysis results. Therefore, the impact of control overhead is taken into account to design S-MUCRP in this paper, and SWIPT is also introduced to further balance the residual energy among nodes.

3. SWIPT-based multihop uneven clustering routing protocol for CRSNs.

3.1. Network model.

Assuming that N_{total} homogeneous CRSNs nodes and P primary users (PUs) are randomly distributed in the circular area with radius R, and the sink is located at the center. The whole area is partitioned into multiple concentric rings (i.e., layers) centered at the sink, and the layer width is set as R_t (the maximum transmission range of CRSNs nodes). According to the Euclidean distance to the sink, layers are numbered from the inside out as layer $1, 2, \ldots, l_{max}$. Assumptions are made as follows: (1) CRSNs nodes cannot move once deployed, and they carry out accurate spectrum sensing, that is, sensing error is ignored. (2) CRSNs nodes can obtain their residual energy and geographical location, and they can also exchange the above information with neighbors on common control channel (CCC). (3) CHs perform perfect data aggregation, to be specific, each CM sends L bits of data to its CH, and the CH receives and aggregates the data with its own into a single packet with size L. Markov ON/OFF model [29] is leveraged to imitate the dynamic behaviors of PUs, and PUs alternate between ON and OFF states whose durations are independent variables. The energy consumption model in [30] is utilized to quantify the node energy consumption of transmitting or receiving information.

3.2. Design details of S-MUCRP.

According to S-MUCRP, CRSNs nodes are enabled to periodically collect and transmit the sensed data to the sink. S-MUCRP determines cluster radii to manage the control information exchange, and intra-cluster SWIPT is also introduced to achieve intra-cluster energy balance. More details are explained in the following subsections.

3.2.1. *Theoretical derivation of cluster radii.*

With the purpose of balancing the energy consumption among CHs, the relationship between cluster radii of neighboring layers are theoretically derived. Together with the cluster radius of the outmost layer which acts as the initial value, cluster radius of each layer is obtained.

(1) Theoretical derivation of the relationship between cluster radii of neighboring layers

The energy consumption of each CH in layer i is composed of the energy consumption of control overhead, intra-cluster data processing and inter-cluster data forwarding, as shown in Equation (2).

$$\begin{aligned} E_{con}CH(i) =& 3L_1 \times \left(E_{elec} + E_{fs} \times R_{ci}^2\right) + 3(N_i - 1)E_{elec} \times L_1 + (N_i - 1)E_{elec} \\ & \times L_2 + N_i \times E_{DA} \times L_2 + L_2 \times \left(E_{elec} + E_{fs} \times d_{CH(i)\to CH(i-1)}^2\right) \\ & + \frac{\sum_{j=i+1}^{l_{max}} \frac{A_j}{S_j} \times \left(2E_{elec} + E_{fs} \times d_{CH(i)\to CH(i-1)}^2\right) \times L_2}{\frac{A_i}{S_i}} \end{aligned} \tag{2}$$

where l_{max} is the maximum number of layers in the whole network; L_1 and L_2 are the size of control packets and data packets, respectively. E_{elec} is the energy consumption of transceiver electronics per bit; E_{fs} is the energy consumption of power amplifier for sending 1 bit of data; E_{DA} is the energy consumption of data aggregation per bit; N_i is the average number of nodes per cluster in layer i; A_i is the area of layer i and $A_i = (2i-1)\pi R_t^2$; S_i is the average area per cluster in layer i and $S_i = \pi R_{ci}^2$, here, R_{ci} is the cluster radius of layer i; $d_{CH(i)\to CH(i-1)}$ is the average distance between CHs in layer i and their relays in layer $i-1$ or the sink, which can be calculated according to Equation (3).

$$d_{CH(i)\to CH(i-1)} = d_{tosink}(CH(i)) - d_{tosink}(CH(i-1)) \tag{3}$$

where $d_{tosink}(CH(i))$ and $d_{tosink}(CH(i-1))$ are the average distance from the CHs in layer i and layer $i-1$ to the sink, respectively. The expression of $d_{tosink}(CH(i))$ is given in Equation (4).

$$d_{tosink}(CH(i)) = \frac{1}{\pi i^2 R_t^2}\int_{\theta=0}^{2\pi}\int_{r=0}^{iR_t} r^2 dr d\theta = \frac{2}{3} iR_t \tag{4}$$

The energy harvested by each CH in layer i is the total energy transferred by its CMs according to their residual energy, as shown in Equation (5).

$$\begin{aligned} E_{har}CH(i) =& (N_i - 1)\times E_{elec}\times L_2 + N_i \times E_{DA}\times L_2 \\ &+ \left(E_{elec} + E_{fs}\times d^2_{CH(i)\to CH(i-1)}\right)\times L_2 \end{aligned} \tag{5}$$

Similarly, the energy harvested by each CH in layer $i+1$ through intra-cluster SWIPT and its energy consumption can also be figured out. Assuming that the initial energy of all CRSNs nodes is identical, in order to balance the residual energy among CHs in different layers, the net energy consumption per CH of layer i $E_{net}CH(i)$, i.e., the variation of residual energy per round, should be calculated. According to Equation (2) and Equation (5), the $E_{net}CH(i)$ can be obtained as below:

$$\begin{aligned} E_{net}CH(i) =& 3\left(E_{elec} + E_{fs}\times R_{ci}^2\right)\times L_1 + 3(N_i - 1)\times E_{elec}\times L_1 \\ &+ \frac{\sum_{j=i+1}^{l_{max}} \frac{A_j}{S_j}\left(2E_{elec} + E_{fs}\times d^2_{CH(i)\to CH(i-1)}\right)\times L_2}{\frac{A_i}{S_i}} \end{aligned} \tag{6}$$

The net energy consumption per CH of layer $i+1$ can be calculated in the similar way, and by equalizing their net energy consumption, the relationship between the cluster radii of layer i and layer $j(j = i+1)$ can be obtained. If $j = l_{max}$, the relationship is shown in Equation (7).

$$R_{c(l_{max}-1)} = \frac{\sqrt{\left(3L_1 m_1 R_{c(l_{max})}\right)^2 + 3L_1 m_1 m_2}}{3L_1 m_1 + \frac{m_2}{R^2_{c(l_{max})}}} \tag{7}$$

where

$$m_1 = E_{fs} + \frac{1}{50}E_{elec} \tag{8}$$

$$m_2 = \frac{\left(18E_{elec}\times L_2 + 4E_{fs}R_t^2\right)(2l_{max} - 1)}{9(2l_{max} - 3)} \tag{9}$$

If $j = l_{max} - 1$, we have:

$$3L_1 m_1 \left(R^2_{c(l_{max}-2)} - R^2_{c(l_{max}-1)}\right) + \left(\frac{2l_{max}-3}{R^2_{c(l_{max}-1)}} + \frac{2l_{max}-1}{R^2_{c(l_{max})}}\right) m_3 \times L_2 \times \frac{R^2_{c(l_{max}-2)}}{2i-1} - \left(\frac{2l_{max}-1}{R^2_{c(l_{max})}}\right) m_3 \times L_2 \times \frac{R^2_{c(l_{max}-1)}}{2i+1} = 0 \quad (10)$$

where

$$m_3 = 2E_{elec} + \frac{4}{9}E_{fs}R_t^2 \quad (11)$$

By substituting Equation (7) into Equation (10), we have

$$R_{c(l_{max}-2)} = \sqrt{\frac{\frac{3L_1 m_1 R^4_{c(l_{max})}\left(3L_1 m_1 R^2_{c(l_{max})} + m_2\right)}{2m_2^2 + 3L_1 m_1 m_2^2 + 3L_1 m_1 R^4_{c(l_{max})}}}{3L_1 m_1 + \left(\frac{(2i-1)\left(3L_1 m_1 + \frac{m_2}{R^2_{c(l_{max})}}\right)^2}{(3L_1 m_1 R_{c(l_{max})})^2 + 3L_1 m_1 m_2} + \frac{2l_{max}-1}{R^2_{c(l_{max})}}\right) \times \frac{m_3 L_2}{2i-1} - \left(\frac{2l_{max}-1}{R^2_{c(l_{max})}}\right) \times \frac{m_3 L_2}{2i+1}}} \quad (12)$$

The relationship between the cluster radii of arbitrary layer and the outmost layer can be acquired. Therefore, the initial value $R_{c(l_{max})}$ should be determined so that the cluster radius of each layer can be derived.

(2) Determining the cluster radius of the outmost layer

The cluster radius of the outmost layer is theoretically derived with the purpose of minimizing the summation of the energy consumption of layer l_{max} and the energy consumption of inner layers for relaying data from layer l_{max}. The energy consumption of layer l_{max} is composed of 2 parts, i.e., the energy consumption of CHs and CMs, as shown in Equation (13) and Equation (14), respectively.

$$\begin{aligned} E_{con}CH(l_{max}) = \Big[&3L_1 \times \left(E_{elec} + E_{fs} \times R^2_{c(l_{max})}\right) + 3(N_{l_{max}} - 1)E_{elec} \\ &\times L_1 + (N_{l_{max}} - 1)E_{elec} \times L_2 + N_{l_{max}} \times E_{DA} \times L_2 \\ &+ L_2 \times \left(E_{elec} + E_{fs} \times d^2_{CH(l_{max}) \to CH(l_{max}-1)}\right)\Big] \times \left(\frac{A_{l_{max}}}{S_{l_{max}}}\right) \end{aligned} \quad (13)$$

$$\begin{aligned} E_{con}CM(l_{max}) = \Big[&3L_1 \times (E_{elec} + E_{fs} \times R^2_{c(l_{max})}) + 2(N_{l_{max}} - 1)E_{elec} \times L_1 + 2L_1 \\ &\times E_{elec} + L_2 \times \left(E_{elec} + E_{fs} \times d^2_{CM \to CH(l_{max})}\right)\Big] \times (N_{l_{max}} - 1) \end{aligned} \quad (14)$$

The total energy consumption of inner layers for relaying the data from layer l_{max} is:

$$E_{relay}(l_{max}) = \sum_{j=0}^{l_{max}-1} \left[\frac{A_{l_{max}}}{S_{l_{max}}} \times L_2 \times \left(2E_{elec} + E_{fs} \times d^2_{CH(j) \to CH(j+1)}\right)\right] \quad (15)$$

where $CH(0)$ is the sink, and $d^2_{CH(0) \to CH(1)}$ denotes the average squared distance from the sink to the CHs in layer 1.

As stated above, the objective is:

$$\begin{aligned} \text{minimize } E_{total}(l_{max}) &= \text{minimize } \{E_{con}CH(l_{max}) + E_{con}CM(l_{max}) + E_{relay}(l_{max})\} \\ &= \text{minimize } \{aR^4_{c(l_{max})} + bR^2_{c(l_{max})} + c\frac{1}{R^2_{c(l_{max})}} + z\} \end{aligned} \quad (16)$$

where a, b, c and z are constants larger than 0, and their expressions are shown in Equation (17) to Equation (20).

$$a = \frac{1}{R_t}\left[E_{fs} \times (3L_1 + L_2) + \frac{2}{R_t} \times L_1 \times E_{elec}\right] \tag{17}$$

$$b = \frac{E_{elec}}{R_t} \times (3L_1 + L_2) - \left[E_{fs} \times (3L_1 + \frac{L_2}{2}) + \frac{2}{R_t} \times L_1 \times E_{elec}\right] \tag{18}$$

$$c = \left[(2l_{max} - 1) \times L_2 \times R_t^2\right] \times \left[\frac{4}{9}l_{max} \times E_{fs} \times R_t^2 + 2(l_{max} - 1) \times E_{elec}\right] \tag{19}$$

$$z = (2l_{max} - 1) \times R_t \times [3L_1 \times (E_{fs} \times R_t + E_{elec}) + L_2 \times (E_{DA} + E_{elec})] - E_{elec}(3L_1 + L_2) \tag{20}$$

By taking the first derivative of Equation (16) with respect to $R^2_{c(l_{max})}$ and setting the result to 0, the cluster radius of layer l_{max} can be derived by taking the square root of the obtained result.

3.2.2. *Energy level function-based cluster formation and multihop route establishment.* Each CRSNs node k determines its geographical location (x_k, y_k), number of available channels $C(k)$ and the Euclidean distance to the sink $d_{tosink}(k)$. Through control information exchange, it can obtain the number of neighbors in the neighboring outer layer which share common available channels and are within R_t $Next(k)$ and number of neighbors in the same layer and within the cluster radius $num(k)$. Based on the above information, node k judges its layer number $l(k)$ and calculates the total energy consumption of processing the data within the cluster radius $E_{neighbor}(k)$ and the energy consumption of forwarding data for outer layers $E_{forward}(k)$ according to Equation (21) to Equation (23).

$$l(k) = \left\lceil \frac{d_{tosink}(k)}{R_t} \right\rceil \tag{21}$$

$$\begin{aligned} E_{neighbor}(k) = &(N_{l(k)} - 1) \times E_{elec} \times L_2 + N_{l(k)} \times E_{DA} \times L_2 \\ &+ \left(E_{elec} + E_{fs} \times d^2_{CH(l(k)) \to CH(l(k)-1)}\right) \times L_2 \end{aligned} \tag{22}$$

$$E_{forward}(k) = \frac{\sum_{j=l(k)+1}^{l_{max}} \frac{A_j}{S_j}}{\frac{A_{l(k)}}{S_{l(k)}}} \times \left(2E_{elec} + E_{fs} \times d^2_{CH(l(k)) \to CH(l(k)-1)}\right) \times L_2 \tag{23}$$

where A_j/S_j is the total number of clusters in layer j. $\sum_{j=l(k)+1}^{l_{max}} \frac{A_j}{S_j}$ is the total number of packets which need to be forwarded by layer $l(k)$.

CRSNs node k calculates its energy level function $ELF(k)$ which is composed of its residual energy $E_{res}(k)$, $E_{neighbor}(k)$ and $E_{forward}(k)$, as shown in Equation (24) below.

$$ELF(k) = \begin{cases} E_{res}(k) - E_{forward}(k) & if\ l(k) = 1 \\ E_{res}(k) - E_{neighbor}(k) - E_{forward}(k) & otherwise \end{cases} \tag{24}$$

Based on the above $ELF(k)$, node k computes its CHs weight $W(k)$ according to Equation (25).

$$W(k) = \begin{cases} [\beta \times ELF(k)]^2 \times [C(k)]^{\frac{1}{3}} \times \left[\dfrac{1}{d_{tosink}(k)}\right]^{\frac{1}{2}} \times \left[\dfrac{1}{Next(k)}\right]^{\frac{1}{2}} \times [num(k)]^{\frac{1}{3}} & if\ l(k) \neq 1 \cap Next(k) \neq 0 \\ 0 & if\ Next(k) = 0 \\ [\beta \times ELF(k)]^2 \times [C(k)]^{\frac{1}{3}} \times \left[\dfrac{1}{d_{tosink}(k)}\right]^{\frac{1}{2}} \times \left[\dfrac{1}{Next(k)}\right]^{\frac{1}{2}} & if\ l(k) = 1 \cap Next(k) \neq 0 \end{cases} \tag{25}$$

where β is the weight coefficient to quantify the impact of $ELF(k)$, and its value is set to 10 in this paper.

The pseudo code of CHs selection is shown in Algorithm 1. Lines 3-11 show that nodes in layer 1 become independent CHs, and other nodes whose residual energy is higher than 0 broadcast CHs weight within cluster radius. Lines 12-40 exhibit how to compete for CHs in the locality. To be specific, each node receives information from neighbors and compares their CHs weight. If its weight is smaller than one of its neighbors, it broadcasts quit message, otherwise it becomes a CH and broadcasts CHs notification message on CCC. The above process is repeated until all nodes become CHs or quit from competition.

Algorithm 1: CHs selection

1.Input: $\boldsymbol{Uncover}=\emptyset$, $\boldsymbol{CH}=\emptyset$, $CHnum=0$, $\boldsymbol{Nei}(k)=\emptyset$, $l(k)$, $W(k)$, $R_{c(l(k))}$, $k=\{1, 2, \ldots, N_{total}\}$.
2.Output: $\boldsymbol{CH}$, $CHnum$.
3.for k=1:1:N_{total}
4.   if $l(k)$==1
5. $\boldsymbol{CH} \leftarrow \boldsymbol{CH} + \{k\}$.
6. $CHnum=CHnum+1$.
7.   else
8. $\boldsymbol{Uncover}=\boldsymbol{Uncover}+\{k\}$.
9.   node k broadcasts CHs weight value $W(k)$ to neighbors.
10.   end
11.end
12.while $\boldsymbol{Uncover} \neq \emptyset$
13.   for k=1:1:N_{total}
14.   if $k \in \boldsymbol{Uncover}$
15.   ch(k)=0. //set initial value.
16. $\boldsymbol{Nei}(k)=\{g | d_{k,g} \leq R_{c(l(k))}$ & $l(k) = l(g)$ & $g \in \boldsymbol{Uncover}\}$.
17.   for s=1:1:N_{total}
18.   if $s \in \boldsymbol{Nei}(k)$ & $W(k) < W(s)$
19.   ch(k)=-1.
20.   break
21.   end
22.   end
23.   if ch(k)==0
24. $\boldsymbol{CH} \leftarrow \boldsymbol{CH} + \{k\}$.
25. $CHnum=CHnum+1$.
26. $\boldsymbol{Uncover} \leftarrow \boldsymbol{Uncover}/\{k\}$.
27.   node k becomes a CH and broadcasts CHs notification message to

neighbors.
28. for s=1:1:N_{total}
29. if $s \in \boldsymbol{Nei}(k)$
30. after receiving the CHs notification message, node s broadcasts quit message.
31. $\boldsymbol{Uncover} \leftarrow \boldsymbol{Uncover}/\{s\}$.
32. end
33. end
34. else
35. Node k broadcasts quit message.
36. $\boldsymbol{Uncover} \leftarrow \boldsymbol{Uncover}/\{k\}$.
37. end
38. end
39. end
40.end

For cluster formation, on receiving CHs notification message, normal nodes will choose the one with the highest weight value as their CH and unicast joining request. The CH receives the request and adds them into its CMs set. Normal nodes which cannot receive any CHs notification message automatically become CHs. The cluster formation is completed when all normal nodes have chosen their CHs, and then all nodes enter into route selection stage.

Nodes in layer 1 can reach the sink through single-hop communication, but restricted by communication range, all other CHs need to select appropriate next-hop relays to assist in data forwarding until the data is transferred to the sink. To be specific, CH k chooses 2 CHs which possess the highest CHs weight values in the next inner 2 layers. If such relays cannot be found, it will select one of its CMs and require it to search for next-hop relay. CHs in layer 2 choose the CHs with the highest CHs weight values in layer 1 as next-hop relays. If such relays cannot be found, they will select one of their CMs and continue to search for next-hop relays.

After cluster construction and route selection, CHs assign time slots to their CMs to schedule their data transmission, and CMs can also obtain the residual energy of their CHs from the schedule information. In this case, they decide whether they will transfer energy and data simultaneously to their CHs. According to the residual energy of CH k $E_{res}(CH(k))$ and the energy consumption of intra-cluster data processing $E_{intra}CH(k)$ (shown in Equation (26)), if $E_{res}(CH(k)) \geq E_{intra}CH(k)$, CMs only transmit data, otherwise they need to transfer the amount of $E_{intra}CM(k)$ energy to CH k for energy compensation.

$$E_{intra}CH(k) = L_2 \times \left[(E_{elec} + E_{DA}) \times N_{l(CH(k))} + E_{fs} \times d^2_{CH(l(CH(k)))\to CH(l(CH(k))-1)}\right] \tag{26}$$

$$E_{intra}CM(k) = \frac{E_{intra}CH(k)}{N_{l(CH(k))} - 1} \times \frac{1}{\eta} \times d^2_{CM\to CH(k)} \tag{27}$$

where η is the energy transfer efficiency. $d^2_{CM\to CH(k)}$ is the average squared distance from CMs to CH k, and its value is $\frac{R^2_{c(l(k))}}{2}$.

If $E_{res}(CM(k)) - E_d(l(CH(k))) \geq E_{intra}CM(k)$, the CM will transfer the amount of $E_{intra}CM(k)$ energy to its CH, and the CH will harvest the amount of $E_{intra}CH(k)/(N_{l(CH(k))} - 1)$ energy. Here, $E_d(l(CH(k)))$ is the energy consumption per data transmission which is shown in Equation (28). If $E_{res}(CM(k)) - E_d(l(CH(k))) < E_{intra}CM(k)$, the CM will transfer data only.

$$E_d(l(CH(k))) = \left(E_{elec} + E_{fs} \times d^2_{CM \to CH(k)}\right) \tag{28}$$

CMs transfer their sensed data together with energy to their CH, and the CH will aggregate the received data with its own and then forward it to the sink.

4. **Simulation results and analysis.** By using Matlab, the number of living nodes, total control overhead, number of effective data gathering nodes and average packet delivery ratio of S-MUCRP are evaluated, and it is compared with the existing clustering routing protocols for CRSNs such as CogLEACH [17], DSAC [18], NSAC [20] and WCM [22] to validate its effectiveness. The detailed simulation parameter settings are shown in Table 2.

Table 2. Simulation parameter settings

Parameters	Values
Network radius (R)	150m
Total number of CRSNs nodes (N_{total})	450
Number of PUs (P)	5
Control packet size (L_1)	100bit
Data packet size (L_2)	1024bit
Energy consumption of data aggregation per bit (E_{DA})	5nJ/bit/packet
Energy consumption of transceiver electronics per bit (E_{elec})	50nJ/bit
Power amplifier coefficient in free-space path loss model (E_{fs})	10pJ/bit/m^2
The maximum transmission range of CRSNs nodes (R_t)	50m

CHs selection, cluster formation, route establishment and data transmission all consume limited node energy, and when their residual energy drops to 0, CRSNs nodes cannot perceive environment or carry out the above operation. Therefore, node energy consumption is an important aspect to evaluate the performance of clustering routing protocols, and it can be analyzed from the number of living nodes which is shown in Figure 1.

As can be seen from Figure 1, the first death node of S-MUCRP appears in round 900, while the first death nodes of DSAC, WCM, NSAC and CogLEACH appear in round 198, 477, 422 and 3422, respectively, which indicates that the node energy consumption of S-MUCRP is relatively low. The reasons can be analyzed as below: (1) As shown in Figure 2, the total control overhead of S-MUCRP is pretty low, and it is more than twice but less than three times the number of living nodes. To be specific, firstly, CRSNs nodes in layer 1 act as independent CHs, and they do not need to broadcast CHs weight, which avoids control information exchange. Secondly, all nodes (except those in layer 1) broadcast CHs weight information to neighbors for CHs selection; Nodes determine whether they can become CHs or not according to the received information. CHs broadcast CHs notification message, and normal nodes which can receive the CHs notification message broadcast quit message. NonCHs nodes request to join the CHs which possess the highest CHs weight values and share common available channels, and corresponding CHs receive the request and record. (2) The control overhead of WCM in CHs selection and cluster formation is about 4 times the number of living nodes. All CRSNs nodes broadcast spectrum sensing results and CHs weight values on CCC for CHs competition; Nodes decide whether they can become CHs by weight comparison in the locality, and CHs broadcast CHs notification message while normal nodes broadcast quit message. NonCHs nodes send joining request to CHs which possess the largest weight value and share common available channels, and CHs deliver their cluster information towards the sink. In DSAC, each CRSNs node is

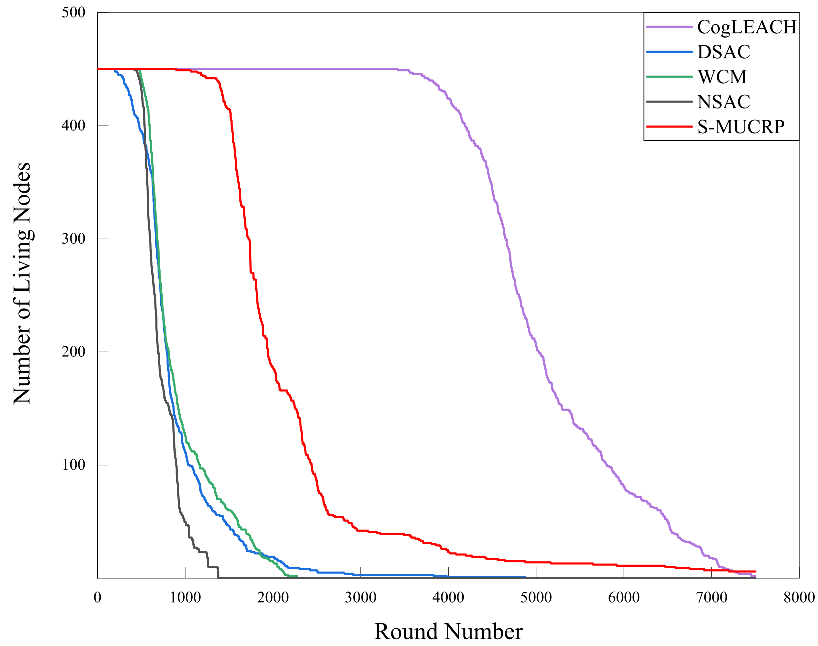

Figure 1. Comparison results of number of living nodes

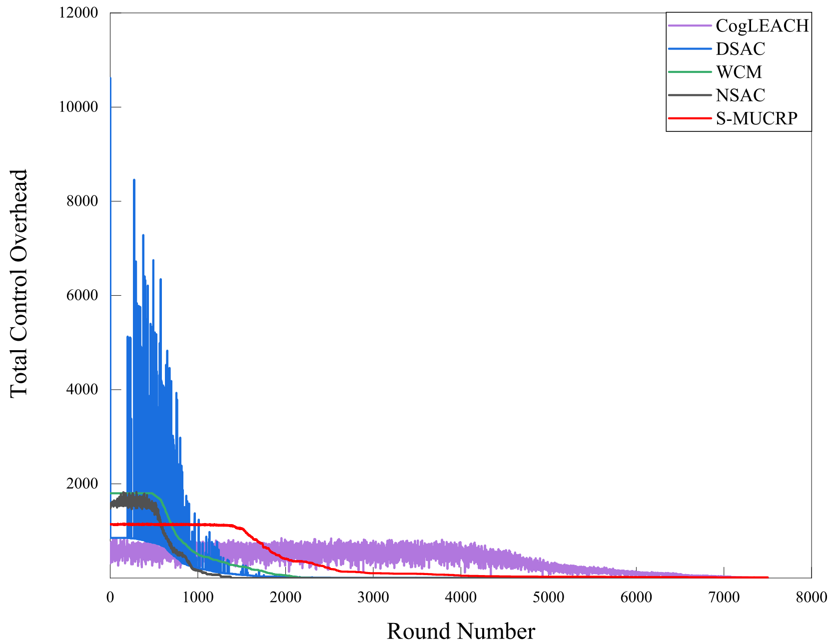

Figure 2. Comparison results of total control overhead

initialized as a CH, and clusters are continuously merged according to common available channels and inter-cluster distance until the optimal number of clusters is achieved. The above process requires extensive control information exchange between CMs and their CHs or between neighboring CHs, which consumes a huge amount of energy. In NSAC, all CRSNs nodes constantly calculate and update their CHs weight values based on residual energy and channel quality. Nodes with the largest weight value in the locality become CHs, and their neighbors join in to form clusters. The above process is repeated until all nodes are clustered. Extensive control information exchange among neighbors is also required which is pretty energy-consuming. As a result, the first death nodes of WCM, DSAC and NSAC appear earlier than S-MUCRP, and their number of living nodes declines rapidly afterwards. The first death node of CogLEACH appears in round 3422, which is later than S-MUCRP. The control overhead of CogLEACH in CHs selection and cluster construction is as low as twice the number of living nodes. To be specific, each CH broadcasts temporary and final CHs notification message while nonCHs nodes send out joining request and confirmation message. Although the node energy consumption of CogLEACH is small, only CHs which can reach the sink directly can deliver their data to the sink. Therefore, the network scalability and surveillance capability are severely restricted.

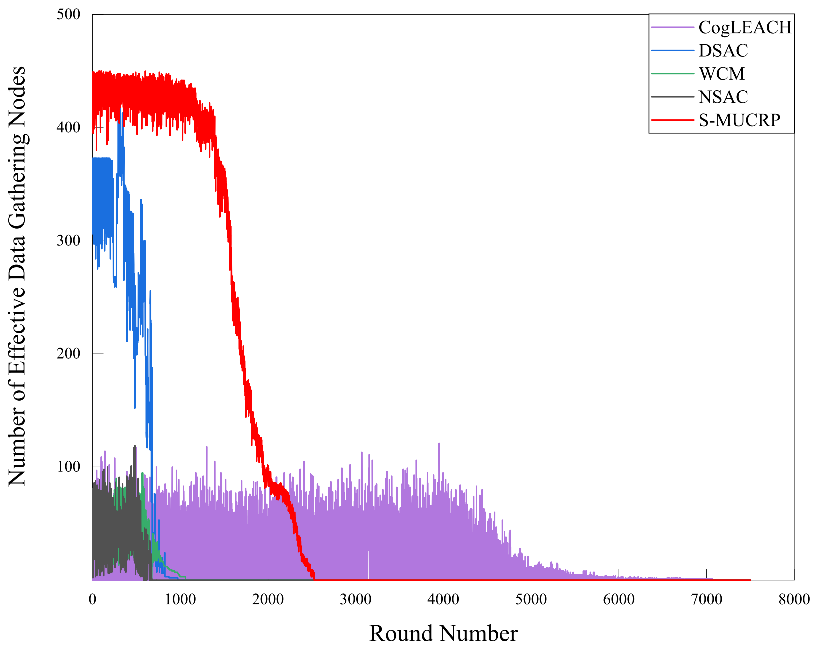

FIGURE 3. Comparison results of number of effective data gathering nodes

Apart from energy consumption, the number of effective data gathering nodes is also an important metric to evaluate the performance of clustering protocols, and it reveals their network surveillance capability. The number of effective data gathering nodes of each protocol is recorded, and the results are shown in Figure 3. DSAC is a multihop clustering routing protocol, and majority of nodes can deliver their data to the sink through multihop routing, therefore, the number of effective data gathering nodes is high at the beginning. However, high control overhead quickly exhausts node residual energy and results in fast decline of the number of living nodes and the number of effective data gathering nodes after round 680. CogLEACH, WCM and NSAC are all single-hop clustering routing protocols for CRSNs, and only nodes which can reach the sink through single-hop communication can effectively deliver their data, correspondingly, the number of effective data gathering nodes is relatively small. In addition, the number of effective data gathering nodes of WCM and NSAC dramatically decreases as the number of living nodes decreases. In S-MUCRP, the channel which is available to the majority of CMs is chosen to transmit data, and other channels will also be determined for CMs if the cluster channel is unavailable. The selected channels are rather stable, and channel reclaim from PUs is rare. All these contribute to more effective data gathering nodes. To further quantify the network surveillance capability of each protocol, the average packet delivery ratio is calculated, and it is defined as the average ratio of effective data gathering nodes

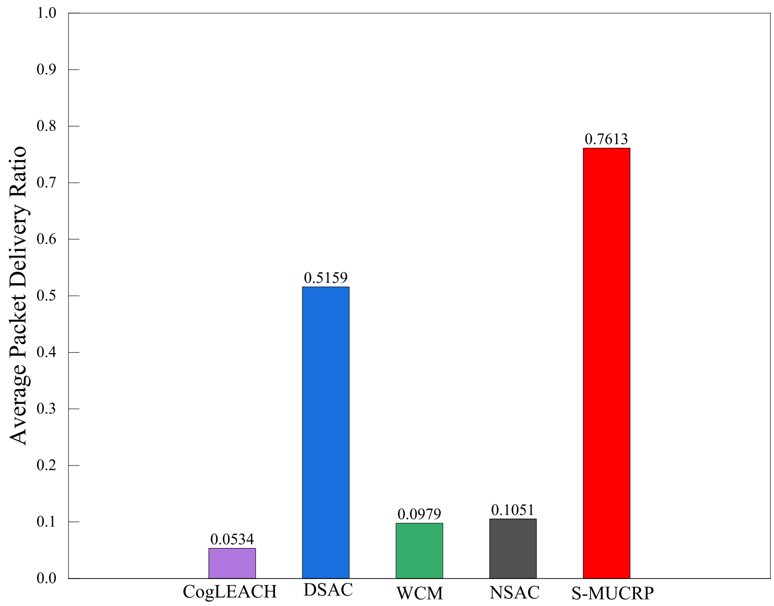

Figure 4. Comparison results of average packet delivery ratio

in all effective rounds. The results are shown in Figure 4. Combining with Figure 3, it can be observed that if their residual energy is sufficient, the effective data gathering nodes of S-MUCRP takes up over 95% of all living nodes, otherwise its number of effective data gathering nodes decreases, but its average packet delivery ratio is still much higher than CogLEACH, DSAC, WCM and NSAC. To be specific, compared with CogLEACH, DSAC, WCM and NSAC, the average packet delivery ratio of S-MUCRP is improved by 70.79%, 24.54%, 66.34% and 65.62%, respectively.

In terms of energy consumption, S-MUCRP can dramatically reduce node energy consumption and prolong the network lifetime. The reasons are summarized as below: (1) Nodes in layer 1 become independent CHs, which can help reduce the amount of control information exchanged for cluster formation. (2) Reasonable configuration of cluster radius can manage the range of control information exchange and dramatically reduce the energy consumption of control overhead. (3) The residual energy of CHs with high energy consumption can be compensated through intra-cluster SWIPT, which can help further balance the residual energy among nodes within the same cluster and postpone their death. In terms of network surveillance capability, S-MUCRP can enhance packet delivery ratio and guarantee powerful network surveillance. The reasons are summarized as follows: (1) Energy level function-based relay selection criterion helps select appropriate relay nodes to establish multihop routes and forward the sensed data towards the sink. (2) Stable available channels are chosen to reduce the collision probability with PUs and achieve effective data delivery. In a word, S-MUCRP gains obvious advantages over existing clustering protocols for CRSNs.

5. **Conclusions.** To achieve effective data delivery in large-scale CRSNs, an SWIPT-based multihop uneven clustering routing protocol S-MUCRP is proposed in this paper. Theoretical derivation is leveraged to obtain the cluster radii of different layers so that the range of control information exchange and corresponding energy consumption can be effectively controlled. Intra-cluster SWIPT is introduced, and CMs decide whether they will transfer energy to their CHs according to their residual energy and the request of their CHs. Simulation results show that S-MUCRP gains obvious advantages in extending

the network lifetime and enhancing the network surveillance capability. To be specific, by reducing the amount of control information exchanged for cluster formation and managing the range of control information exchange, the first death node of S-MUCRP is postponed to round 900 which is only earlier than CogLEACH. Through reasonable relay selection and channel selection, the sensed data can be delivered towards the sink effectively, and the number of effective data gathering nodes accounts for over 95% of all living nodes when the remaining energy is sufficient. On average the packet delivery ratio of S-MUCRP is over 24% higher than DSAC and over 65% higher than other competing protocols. In our future work, we plan to incorporate radio frequency energy harvesting and inter-cluster SWIPT to further compensate for limited node energy and prolong the network lifetime.

REFERENCES

[1] T. Y. Wu, L. Yang, Z. Y. Lee, S. C. Chu, S. Kumari, and S. Kumar, "A provably secure three-factor authentication protocol for wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2021, 5537018, 2021.

[2] X. A. Li, S. S. Liu, S. Kumari, and C. M. Chen, "PSAP-WSN: A provably secure authentication protocol for 5G-based wireless sensor networks," *CMES-Computer Modeling in Engineering & Sciences*, vol. 135, no. 1, pp. 711–732, 2023.

[3] J. N. Chen, Y. P. Zhou, Z. J. Huang, T. Y. Wu, F. M. Zou, and R. Tso, "An efficient aggregate signature scheme for healthcare wireless sensor networks," *Journal of Network Intelligence*, vol. 6, no. 1, pp. 1–15, 2021.

[4] K. Renuka, S. Kumar, S. Kumari, and C. M. Chen, "Cryptanalysis and improvement of a privacy-preserving three-Factor authentication protocol for wireless sensor networks," *Sensors*, vol. 19, no. 21, 4625, 2019.

[5] J. H. Wang, and S. Li, "ECE: a novel performance evaluation metric for clustering protocols in cognitive radio sensor networks," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 2078–2079, 2021.

[6] Z. X. Liu, M. Y. Zhao, Y. Z. Yuan, and X. P. Guan, "Subchannel and resource allocation in cognitive radio sensor network with wireless energy harvesting," *Computer Networks*, vol. 167, pp. 1–10, 2020.

[7] J. N. Chen, F. M. Zou, T. Y. Wu, and Y. p. Zhou, "A new certificate-based aggregate signature scheme for wireless sensor networks," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 5, pp. 1264–1280, 2018.

[8] C. M. Chen, Y. H. Lin, Y. C. Lin, and H. M. Sun, "RCDA: recoverable concealed data aggregation for data integrity in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 727–734, 2011.

[9] J. Guo, and H. Jafarkhani, "Movement-efficient sensor deployment in wireless sensor networks with limited communication range," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3469–3484, 2019.

[10] R. Prajapat, R. N. Yadav, and R. Misra, "Energy efficient k-hop clustering in cognitive radio sensor network for internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13593–13607, 2021.

[11] A. Shahraki, A. Taherkordi, O. Haugen, and F. Eliassen, "A survey and future directions on clustering: from WSNs to IoT and modern networking paradigms," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 2242–2274, 2021.

[12] J. H. Wang, and W. X. Shi, "Survey on cluster-based routing protocols for cognitive radio sensor networks," *Journal on Communications*, vol. 39, no. 11, pp. 156–169, 2018.

[13] L. R. Varshney, "Transporting information and energy simultaneously," *2008 IEEE International Symposium on Information Theory*, Toronto, Canada, pp. 1612–1616, 2008.

[14] D. M. S. Bhatti, N. Saeed, and H. Nam, "Fuzzy c-means clustering and energy efficient cluster head selection for cooperative sensor network," *Sensors*, vol. 16, no. 9, 1459, 2016.

[15] J. H. Wang, and S. Li, "Ions motion optimization-based clustering routing protocol for cognitive radio sensor network," *IEEE Access*, vol. 8, pp. 187766–187782, 2020.

[16] A. Latiwesh, and D. Y. Qiu, "Energy efficient spectrum aware clustering for cognitive sensor networks: CogLEACH-C," *2015 10th International Conference on Communications and Networking in China (ChinaCom)*, Shanghai, China, pp. 512–520, 2015.

[17] R. M. Eletreby, H. M. Elsayed, and M. M. Khairy, "CogLEACH: a spectrum aware clustering protocol for cognitive radio sensor networks," *2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, Oulu, Finland, pp. 179–184, 2014.

[18] H. Z. Zhang, Z. Y. Zhang, H. Y. Dai, R. Yin, andX. M. Chen, "Distributed spectrum-aware clustering in cognitive radio sensor networks," *2011 IEEE Global Telecommunications Conference*, Houston, USA, pp. 1–6, 2011.

[19] G. A. Shah, F. Alagoz, E. A. Fedal, and O. B. Akan, "spectrum-aware clustering for efficient multimedia routing in cognitive radio sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 7, pp. 3369–3380, 2014.

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

[21] R. N. Yadav, R. Misra, and D. Saini, "Energy aware cluster based routing protocol over distributed cognitive radio sensor network," *Computer Communications*, vol. 129, pp. 54–66, 2018.

[22] T. J. Wang, X. J. Guan, X. L. Wan, H. Shen, and X. M. Zhu, "A spectrum-aware clustering algorithm based on weighted clustering metric in cognitive radio sensor networks," *IEEE Access*, vol. 7, pp. 109555–109565, 2019.

[23] E. R. Pei, H. Z. Han, Z. H. Sun, B. Shen, and T. Q. Zhang, "LEAUCH: low-energy adaptive uneven clustering hierarchy for cognitive radio sensor network," *Eurasip Journal on Wireless Communication and Networking*, vol. 2015, no. 1, pp. 1–8, 2015.

[24] C. Q. Wang, and S. B. Wang, "Research on uneven clustering APTEEN in CWSN based on ant colony algorithm," *IEEE Access*, vol. 7, pp. 163654–163664, 2019.

[25] T. Stephan, F. Al-Turjman, J. K. Suresh, and B. Balusamy, "Energy and spectrum aware unequal clustering with deep learning based primary user classification in cognitive radio sensor networks," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 11, pp. 3261–3294, 2021.

[26] M. C. Zhang, R. J. Zheng, Y. Li, Q. T. Wu, and L. Song, "R-bUCRP: A novel reputation-based uneven clustering routing protocol for cognitive wireless sensor networks," *Journal of Sensors*, vol. 2016, 5986265, 2016.

[27] H. H. Choi, and J. R. Lee, "Energy-neutral operation based on simultaneous wireless information and power transfer for wireless powered sensor networks," *Energies*, vol. 12, no. 20, 3823, 2019.

[28] M. M. Afsar, and M. Younis, "A load-balanced cross-layer design for energy-harvesting sensor networks," *Journal of Network and Computer Applications*, vol. 145, 102390, 2019.

[29] A. Motamedi, and A. Bahai, "MAC protocol design for spectrum-agile wireless networks: stochastic control approach," *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Dublin, Ireland, pp. 448–451, 2007.

[30] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transaction on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.