Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **8**, Number **3**, August **2023**

A Simultaneous Wireless Information and Power Transfer-based Multi-hop Clustering Routing Protocol for RF Energy Harvesting-Cognitive Radio Sensor Networks

Ji-Hong Wang[1,2*]

[1]Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology,
Ministry of Education
Northeast Electric Power University
[2]School of Electrical Engineering
Northeast Electric Power University
169 Changchun Road, Jilin, 132012, China
wangjihong07@126.com

Xi-Tong Su

School of Electrical Engineering
Northeast Electric Power University
169 Changchun Road, Jilin, 132012, China
sxt000725@126.com

Yi-Yang Ge

School of Electrical Engineering
Northeast Electric Power University
169 Changchun Road, Jilin, 132012, China
gyy0610@126.com

Yeh-Cheng Chen

Department of Computer Science
University of California
Davis, CA, 001313, USA
ycch@ucdavis.edu

*Corresponding author: Ji-Hong Wang

Received March 14, 2023, revised May 7, 2023, accepted June 24, 2023.

Abstract. *Radio frequency energy harvesting (RF EH) and clustering are promising solutions to solve the energy constraint problem faced by cognitive radio sensor networks (CRSNs). However, study on clustering routing protocol design for RF EH-CRSNs is still in the initial stage, and there lacks of solutions for inter-cluster routing problem, which restricts the network scalability. In addition, unbalanced harvested energy and energy consumption may result in premature death of nodes in key position, which will significantly reduce the network performance. In this paper, a simultaneous wireless information and power transfer (SWIPT)-based multi-hop clustering routing protocol (SMCRP) is proposed for RF EH-CRSNs to well balance the residual energy distribution among nodes and postpone node death. Apart from the RF energy harvested from dedicated or ambient energy sources, the energy obtained through intra-cluster and inter-cluster SWIPT is also considered to facilitate high-quality cluster heads selection, cluster formation and inter-cluster route establishment. Power splitting ratio is derived based on the residual energy of CRSNs nodes and their energy consumption to enable effective implementation of SWIPT in CRSNs. Simulation results show that SMCRP is effective in balancing the residual energy distribution among nodes and improving node sustainability. To be specific, compared with current clustering protocols, the network lifetime is extended by 64.07% while intra-cluster energy balance degree and network energy balance degree are improved by 15.1% and 7.82% , respectively.*
Keywords: Cognitive radio sensor networks, Clustering, RF energy harvesting, Simultaneous wireless information and power transfer

---

1. **Introduction.** As one of the most important underlying technologies of Internet of Things (IoT), wireless sensor networks (WSNs) provide indispensable basis for data collection and analysis in IoT [1]. WSNs are composed of a large number of sensor nodes which are distributed in space and establish the network topology by leveraging wireless communication technology [2, 3]. WSNs possess the advantages of fast deployment, flexible communication and dynamic reconfiguration, and they are widely applied in field such as natural environment monitoring [4], medical and health care [5] and space exploration [6]. However, WSNs nodes are usually small in size, and the energy stored in battery is limited, which leads to serious energy constraint problem. In addition, spectrum shortage is also one of the challenges which need to be solved urgently in WSNs.

Cognitive radio sensor networks (CRSNs) are new network paradigms for high spectral efficiency, and they solve the spectrum shortage problem in traditional WSNs by smartly incorporating cognitive radio (CR) technology [7]. However, CRSNs nodes still adopt small-size and capacity-constrained battery as their energy source, and executing CR functions will further consume their limited battery energy and result in more severe energy constraint problem [8].

To reduce the energy consumption of CRSNs nodes, researchers have proposed a variety of strategies, such as energy management and scheduling [9], optimizing node distribution and route selection [10], task offloading and re-deployment [11] and so on. However, the above methods only consider about how to reduce the energy consumption, but no extra energy is supplemented. Therefore, the energy consumption problem cannot be fundamentally solved. In addition, the node distribution optimization and route selection scheme still needs to transmit a huge amount of redundant data, which restricts its energy-saving performance. In order to further solve the energy constraint problem of CRSNs nodes, radio frequency energy harvesting (RF EH) and clustering are introduced into CRSNs. RF EH enables CRSNs nodes to harvest energy from dedicated or ambient sources to compensate for limited battery energy; Clustering reduces data redundancy and the corresponding energy consumption through data fusion and aggregation [12, 13].

Therefore, cluster-based RF EH-CRSNs have great potentials in improving energy and spectral efficiency, which are particularly appealing to future IoT.

In order to achieve a more balanced residual energy distribution among CRSNs nodes and postpone their death, simultaneous wireless information and power transfer (SWIPT) technology [14] is incorporated into CRSNs, and efficient clustering routing protocols are required to form clusters and gather useful information in such network. However, current clustering routing protocols still have the following defects: (1) Existing clustering routing protocols are mainly put forward for nonEH-CRSNs. Energy is drained from the small-size and capacity-constrained battery by spectrum sensing, cluster heads (CHs) selection, cluster formation and other operations. Without extra energy supplement, the node survival time is severely restricted. (2) Research on RF EH-CRSNs does not provide solutions for inter-cluster routing problem, which results in severely-constrained network scalability. (3) Current SWIPT-based clustering routing protocols are unsuitable for RF EH-CRSNs, because they are usually devised for WSNs working on the unlicensed frequency band and ignore dynamic channel availability. On the basis of our previous work in [15], a SWIPT-based multi-hop clustering routing protocol (SMCRP) is proposed for RF EH-CRSNs in this paper to care more about the nodes with high energy consumption. The innovations of this paper are summarized as follows:
(1) Apart from the RF energy harvested from dedicated or ambient energy sources, the energy obtained through intra-cluster and inter-cluster SWIPT is also taken into consideration to choose high-quality CHs, form clusters and establish effective inter-cluster routes, which can promote the energy efficiency dramatically.
(2) In order to enable effective implementation of SWIPT, aiming at balancing the residual energy among nodes, power splitting (PS) ratio ρ is derived based on the residual energy of CRSNs nodes and their energy consumption. In this case, nodes with high energy consumption are enabled to accurately decode information and obtain energy supplement simultaneously.
(3) Intra-cluster and network normalized energy balancing indicators are defined to quantitatively evaluate the capability of clustering protocols in balancing the residual energy among nodes within different ranges. It is verified by simulation results that SMCRP has obvious advantages when compared with other competitive protocols in balancing the residual energy among nodes and prolonging the network lifespan.

2. **Related works.** Existing clustering routing protocols are mainly put forward for nonEH-CRSNs. Energy is drained from the small-size and capacity-constrained battery by spectrum sensing, CHs selection, cluster formation and other operations, but no extra energy is compensated. Depending on whether there exists a central entity, these clustering routing protocols can be further classified into centralized protocols and distributed protocols.

CogLEACH-C [16], Fuzzy C-means [17] and IMOCRP [18] are representatives of centralized clustering protocols for nonEH-CRSNs. For example, by leveraging artificial intelligence algorithm, the sink determines the most suitable CHs and the cluster membership based on all information received from CRSNs nodes in IMOCRP. In this case, the optimal number of clusters is automatically determined without complicated theoretical derivation. Centralized clustering protocols are characterized by easy implementation and low control overhead, but it is required that all nodes can reach the sink directly. As a result, the network scalability is severely constrained, and a single point of failure may occur [19].

In contrast, distributed clustering routing protocols such as CogLEACH [20], NSAC [21], WCM [22], DSAC [23], EACRP [24] and SCEEM [25] rely on local information exchange for CHs selection, cluster formation and even inter-cluster route establishment. They can conquer the limitations of centralized clustering protocols. However, multi-hop clustering protocols organize the clustering architecture through extensive control information exchange which consumes too much energy. For example, DSAC first treats all nodes as CHs and then merges adjoining clusters in accordance with the Euclidean distance between clusters and their common channels until the number of clusters which can cause the minimum network energy dissipation is reached. In order to judge whether merging conditions are satisfied, a huge amount of control information is exchanged in the locality, which quickly drains node battery and negatively affects the packet delivery ratio.

EH is applied to power CRSNs nodes and break the performance bottleneck imposed by capacity-constrained battery. According to previous studies, in contrast to the unstable energy provided by natural sources, RF energy source is relatively stable, and the amount of energy obtained by CRSNs nodes can be predicted. Therefore, studies on RF EH gradually increase. Researchers put forward many strategies, such as channel pairing strategy in [26] and new resource allocation strategy in [27], but they cannot be applied to multi-hop RF EH-CRSNs. In our previous work [15], a RF EH-based multi-hop clustering routing protocol (RFMCRP) is proposed to overcome the limitations of current research results. However, EH is position-dependent, and energy consumption is role-dependent. This may result in unbalanced energy distribution among nodes, and dead nodes in key position will significantly reduce the network performance. In order to further balance the residual energy among RF EH-CRSNs nodes and improve their sustainability, SWIPT technology is introduced to improve RFMCRP in this paper. Research on SWIPT-based clustering routing protocol design is still in its infancy, and the research results mainly focus on WSNs such as SWIPT-ENO scheme [28] and SWIPT-based cross-layer scheme CREST [29]. They consider the static spectrum usage and can only be used in WSNs. Therefore, a SWIPT-based clustering routing protocol which is exclusively designed for RF EH-CRSNs is urgently needed to further balance the energy distribution among nodes and prolong the network lifetime.

3. SWIPT-based multi-hop clustering routing protocol design for RF EH-CRSNs.

3.1. System model.

The proposed RF EH-CRSNs are composed of N homogeneous CRSNs nodes which are stochastically and uniformly distributed in the circular monitoring area with radius R. They can opportunistically access the licensed channels authorized to P primary users (PUs) in the same area. The sink located at the center acts as the gateway node towards the network manager. In order to minimize the number of transmission hops and the delay introduced by multi-hop data relaying, the monitoring area is uniformly divided into l_m layers, and CRSNs nodes in the same layer compete for CHs and eventually form clusters according to SMCRP. Here, $l_m = \lceil R/d_n \rceil$, and d_n is the maximum communication range of CRSNs nodes which is also the width of each layer. The layer closest to the sink is numbered as layer 1, i.e., the layer number is a non-decreasing function with respect to the Euclidean distance to the sink. CRSNs node i determines its layer number $l(i)$ according to $l(i) = \lceil d_{i,sink}/d_n \rceil$ based on its Euclidean distance to the sink $d_{i,sink}$. Semi Markov ON/OFF model is leveraged to imitate the dynamic behaviors of PUs on licensed channels, i.e., PUs alternate between ON and OFF states, and the time duration of each state is an independent variable [8]. CRSNs node perceives the occupancy states of licensed channels through spectrum sensing, and sensing error is neglected [30]. If the licensed channel is perceived as idle (in OFF state), it can be opportunistically leveraged by CRSNs nodes for data transmission, otherwise the busy PU serves as an ambient energy source, and CRSNs nodes can harvest RF energy from its emitted signal.

CRSNs nodes compare the gross amount of RF energy which they can harvest from the dedicated energy source (the sink) with that from ambient energy sources (PUs in ON state) within EH duration T, and then they choose to harvest energy from the more powerful source and store the harvested energy into battery for future use. In order to further compensate for nodes with high energy dissipation and balance the residual energy among nodes, SWIPT technology is introduced into RF EH-CRSNs. At present, there are 2 kinds of SWIPT receiver architectures, i.e., time-switching (TS) and PS architectures [31]. Compared with TS architecture, PS architecture is widely researched, and it can achieve a better balance between the information transmission rate and energy conversion [32]. Therefore, PS architecture is adopted in this paper, and the received RF power is divided into 2 parts for information decoding and EH, respectively. PS ratio ρ is defined to represent the proportion of the received RF power used for information decoding, and the remainder ($1-\rho$) is leveraged for EH.

Various kinds of operations consume limited node energy, such as spectrum sensing, transmitting and receiving information. In this paper, the typical model in [33] is used to quantify the energy dissipation of CRSNs nodes. To be specific, when a source node sends L bits of information to its destination d meters away, its energy dissipation is calculated according to Equation (1), and the energy dissipation of receiving the information is shown in Equation (2).

$$E_{Tx}(L,d) = \begin{cases} (E_{elec} + E_{fs} \times d^2) \times L & if\ d \le d_0 \\ (E_{elec} + E_{mp} \times d^4) \times L & otherwise \end{cases} \tag{1}$$

$$E_{Rx}(L) = E_{elec} \times L \tag{2}$$

where E_{elec} is the energy consumption of transceiver electronics per bit. E_{fs} and E_{mp} denote the power amplifier coefficients in free-space and multi-path loss models, respectively. d_0 is the distance threshold and d_0=87.7m. If $d \le d_0$, free-space path loss model is applied to quantify the signal attenuation which is positively proportional to d^2, otherwise multi-path loss model is utilized.

3.2. **Design details of SMCRP protocol.** Aiming at achieving a more balanced residual energy distribution among nodes and enabling a longer network lifetime, SMCRP protocol is proposed on the basis of our previous work RFMCRP [15]. The operation process of SMCRP consists of 5 stages, i.e., spectrum sensing, EH, cluster formation, route selection and data transmission. Its pseudo-code is shown in Algorithm 1 below.

Algorithm 1: SMCRP Protocol
1. // Stage 1: spectrum sensing
2. Each CRSNs node perceives the occupancy states of the licensed channels and determines its energy source.
3. // Stage 2: EH
4. Each CRSNs node harvests energy from the selected source within time duration T.
5. // Stage 3: cluster formation
6. // CHs selection
7. Each running node broadcasts its information table and receives information from neighbors.

8. for i=1:1:N
9. if $l(i) \sim= 1$
10. **CompeteSet**=**CompeteSet** + $\{i\}$.
11. Node i calculates and broadcasts $S\_ELF(i)$ and $S\_Compt(i)$ values.
12. end
13. end
14. for i=1:1:N
15. if $i \in$ **CompeteSet**
16. ch=1.
17. for j=1:1:N
18. if $d_{i,j} \leq d_n$ && $j \in$ **CompeteSet** && $l(j) == l(i)$
19. if $S\_Compt(i) < S\_Compt(j)$
20. ch=0.
21. break
22. end
23. end
24. end
25. if ch == 1
26. Node i becomes a CH and broadcasts CHs notification message on CCC.
27. for j=1:1:N
28. if $d_{i,j} \leq d_n$ && $j \in$ **CompeteSet** && $l(j) == l(i)$
29. Node j broadcasts quit message on CCC.
30. **CompeteSet**=**CompeteSet** $- \{j\}$.
31. end
32. end
33. else
34. Node i broadcasts quit message on CCC.
35. end
36. **CompeteSet**=**CompeteSet** $- \{i\}$.
37. end
38. end
39. // cluster construction
40. Each nonCHs node j temporarily joins CH i which shares common available channels with it and possesses the smallest D(CH i).
41. CH i calculates channel levels and selects its cluster channel $Channel$(CH i).
42. if $Channel$(CH i) $\in \boldsymbol{C}(j)$
43. Node j becomes a final CM of CH i.
44. else
45. Node j selects another CH or becomes an independent CH.
46. end
47. // Stage 4: route selection
48. Node s which requires packet forwarding selects its next-hop relay from candidate relay node set according to Equation (7).
49. The next-hop relay selection process is repeated until a complete route towards the sink is established or any available route cannot be found.
50. // Stage 5: data transmission
51. // intra-cluster data transmission
52. for j=1:1:K
53. if $E_{residual}$(CH j) $< E_{exp}$(CH j)
54. Edge CMs in cluster j transfer energy and information to CH j according to

Equation (14).
    else
        Edge CMs transfer their excessive portion of energy to CH j if they have.
    end
end
// inter-cluster data transmission
for any source node s and relay node q on the established route
    if $E_{residual}(q) < E_{exp}(q)$
        Node s will transfer energy and information to q according to Equation (16).
    else
        edge CH s will transfer its excessive portion of energy to q if it has.
    end
end

(1) Spectrum sensing and EH stages

CRSNs node i whose residual energy is above 0 performs individual spectrum sensing. The available channel list is determined based on the sensing results, and the idle rate of these available channels is calculated according to the historical data. More details can be found in [15]. In addition, based on linear EH model whose conversion efficiency is α, CRSNs node i evaluates the amount of energy which can be harvested from the sink and from the ambient busy PUs, respectively, and then the source from which the maximum amount of energy can be harvested is chosen. Within the subsequent time duration T, node i harvests energy from the RF signal emitted by its selected source.

(2) Cluster formation stage

In cluster formation stage, how to compete for CHs and how a normal CRSNs node determines its CH need to be solved. In order to compete for CHs, CRSNs nodes broadcast control information twice on common control channel (CCC). Firstly, running node i broadcasts its information table which includes the node ID, geographical location, Euclidean distance to the sink $d_{i,sink}$, residual energy $E_{residual}(i)$, amount of harvested energy $E_{Harv}(i)$, available channel list $\boldsymbol{C}(i)$ and idle rate of available channels. As CRSNs nodes can transfer information and RF energy simultaneously through intra-cluster SWIPT, the maximum amount of energy each neighbor can provide needs to be involved in CHs competition. On receiving all information from neighbors, node i calculates its SWIPT-based energy level function $S\_ELF(i)$ as below:

$$\begin{aligned} S\_ELF(i) =& E_{residual}(i) + E_{Harv}(i) + \sum_{k=1}^{Neigh(i)} E_{maxEH}(k) \\ & - \left[Neigh(i) \times E_{elec} \times L + (2 \times E_{elec} + E_{fs} \times \frac{d_n^2}{2}) \times L \times \frac{l_m^2 - l(i)^2}{2 \times l(i) - 1} \right] \end{aligned} \tag{3}$$

where $E_{residual}(i)$ is the residual energy of node i; $E_{Harv}(i)$ is its amount of harvested energy; $Neigh(i)$ is the number of running neighbors of node i; $E_{maxEH}(k)$ denotes the maximum amount of energy the k^{th} neighbor can provide. Correspondingly, the third item on the right side of Equation (3) is the maximum amount of energy can be obtained from neighbors through intra-cluster SWIPT. The first 3 items on the right side of Equation (3) quantify the energy potential of node i, and the fourth item denotes the predicted energy dissipation of node i per round. Based on $S\_ELF(i)$, node i calculates its CHs competition value $S\_Compt(i)$ according to Equation (4).

$$S\_Compt(i) = \frac{S\_ELF(i)}{\left(2 \times E_{elec} + E_{fs} \times \frac{d_n^2}{2}\right) \times L} \times comchannel(i) \times Neigh(i) \times \frac{1}{d_{i,sink}} \quad (4)$$

where the denominator of the first item is the average energy consumption of relaying 1 data packet, and the relay capability of node i is quantified by the first item. $comchannel(i)$ is the average number of available channels shared by node i and its neighbors. Nodes with more available channels, more running neighbors and closer to the sink are more likely to become CHs. Secondly, nonlayer 1 nodes broadcast their energy level function values and CHs competition values to compete for CHs. The detailed CHs competition process is shown in lines 6-38 of Algorithm 1. Figure 1 exhibits how SMCRP selects CHs and next-hop relays. To be specific, the solid circle centered at a node and with radius d_n represents its coverage area, and the nodes within the coverage area constitute its neighbor set. The numbers below each node represent its set of available channels. For example, the red circle represents the coverage area of node i. Its available channel set includes channels 1, 2 and 3, and its neighbor set is $\{j,k,p\}$. Take node i as an example to explain the CHs selection of SMCRP protocol: At the beginning, node i broadcasts its own information, $S\_ELF(i)$ and $S\_Compt(i)$ within d_n and receives the information broadcast by neighbors. Then, node i compares its own weight value $S\_Compt(i)$ with its neighbors: if $S\_Compt(i)>S\_Compt(j)$ & $S\_Compt(i)>S\_Compt(k)$ & $S\_Compt(i)>S\_Compt(p)$, node i becomes a CH and broadcasts CHs notification message. Nodes j, k and p receive the message and broadcast quit messages. Node q receives the quit message from p and deletes it from the CHs competition list. If $S\_Compt(q)>S\_Compt(m)$, node q becomes a CH and broadcasts CHs notification message, and node m receives the CHs notification message and broadcasts a quit message. If $S\_Compt(j)>S\_Compt(i)$ or $S\_Compt(k)>S\_Compt(i)$ or $S\_Compt(p)>S\_Compt(i)$, node i broadcasts a quit message, and nodes j, k and p receive the message and delete i from their CHs competition list. If nodes j, k and p all have neighbors with higher weight, nodes j, k and p broadcast quit messages. In this case, node i becomes an independent CH.

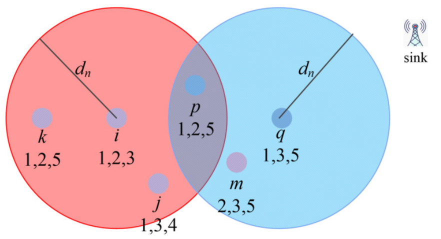

FIGURE 1. CHs selection and route selection of SMCRP.

CHs broadcast CHs notification messages on CCC. NonCHs node j chooses CH i which shares common available channels and with the smallest $D(\text{CH } i)$ (shown in Equation (5)) in the communication range as its CH and sends joining request. CH i receives the request and stores j into its temporary cluster members (CMs) list. Based on the information of all temporary CMs, channel levels are figured out according to Equation (6), and the channel with the highest level is determined as cluster channel. CH i broadcasts the cluster channel information on CCC and deletes the nodes with unavailable channel from the temporary CMs list. Other temporary CMs become final CMs automatically. On receiving the cluster channel information, if it is unavailable at normal CMs, these CMs continue to select CHs and become final CMs by sending joining request. Nodes which cannot receive any cluster channel information become independent CHs.

$$D(\text{CH } i) = d^2_{j,\text{CH } i} + d^2_{\text{CH } i,sink} \tag{5}$$

$$CL_{(\text{CH } i)}(c) = \left[\sum_{k=1}^{Num(\text{CH } i)} Channel_k(c)\right] \times P_{idle(\text{CH } i)}(c) \tag{6}$$

where $Num(\text{CH } i)$ is the number of temporary CMs, and the first item on the right side of Equation (6) denotes the number of temporary CMs on channel c; $P_{idle(\text{CH } i)}(c)$ is the idle rate of channel c at CH i. If there are multiple channels with the same highest level, random selection is leveraged to break the tie. The cluster channel is denoted by $Channel(\text{CH } i)$, and other channels are recorded as backup channels.

(3) Route selection stage

A source node s which cannot reach the sink through single-hop communication selects its next-hop relay with the highest relaying competition value from its candidate relay node set. The candidate relay node set of s is composed of nodes which satisfy the following conditions: ① within the communication range of s; ② share common available channels with s; ③ closer to the sink than s. The relaying competition value of candidate relay q is calculated according to Equation (7).

$$S\_Relaycompt(q) = \frac{1}{|\boldsymbol{c}_{s,q}|} \times \sum_{c=1}^{|\boldsymbol{c}_{s,q}|} \frac{P_{idle(s)}(c) + P_{idle(q)}(c)}{2} \times S\_ELF(q) \times \frac{1}{D(q)} \tag{7}$$

where $P_{idle(s)}(c)$ and $P_{idle(q)}(c)$ are the idle rate of channel c at nodes s and q, respectively. $|\boldsymbol{c}_{s,q}|$ is the number of available channels shared by s and q. $D(q)$ is the distance index calculated by Equation (5). The route selection continues until a whole route is built or any available next-hop relay cannot be found. After cluster formation and route selection, nodes enter into data transmission stage.

Take CH i as an example to explain the route selection of SMCRP protocol. If CH i cannot reach the sink through single-hop communication, it will select a relay node to realize successful data delivery. At the beginning, CH i constructs its candidate relay node set according to the above 3 conditions. In this case, the candidate relay node set of CH i is $\{j,p\}$. CH i calculates $S\_Relaycompt(p)$ and $S\_Relaycompt(j)$ and compares them. If $S\_Relaycompt(p) > S\_Relaycompt(j)$, node p becomes the next-hop relay of CH i; otherwise, node j becomes the next-hop relay. The next-hop relay repeats the above process until the entire route is bulit from CH i to the sink or any available relay nodes cannot be found.

(4) Data transmission stage

The data transmission stage is composed of intra-cluster data transmission (lines 51-58) and inter-cluster data transmission sub-stages (lines 59-66). As intra-cluster and inter-cluster SWIPT are leveraged to transfer information and energy in parallel, the PS ratio ρ should be determined first. According to its definition, $\rho = P_{r\_thresh}/P_r$. Here, P_{r_thresh} is the minimum received power threshold to guarantee successful information decoding, and P_r is the total received signal power. According to the free-space path loss model and multi-path loss model, P_{r_thresh} can be denoted by:

$$P_{r\_thresh} = \begin{cases} \dfrac{P_t \times G_t \times G_r \times \lambda^2}{16\pi^2 \times d^2} & if\ d \le d_0 \\ \dfrac{P_t \times G_t \times G_r \times h_t^2 \times h_r^2}{d^4} & otherwise \end{cases} \tag{8}$$

where P_t is the transmit power. The gains of the transmit antenna and the receiving antenna are denoted by G_t and G_r, respectively. λ is the carrier wavelength. h_t and h_r represent the height of the transmit antenna and the receiving antenna, respectively. d is the Euclidean distance between the transmitter and the receiver. The energy consumption of sending 1 bit of data $E_{TX}(1, d)$ is equal to the product of the transmit power P_t and the reciprocal of the information transmission rate R_b. Combining with Equation (1), we have:

$$P_t = \begin{cases} E_{fs} \times R_b \times d^2 & if\ d \le d_0 \\ E_{mp} \times R_b \times d^4 & otherwise \end{cases} \tag{9}$$

By substituting Equation (9) into Equation (8), P_{r_thresh} is:

$$P_{r\_thresh} = \begin{cases} \dfrac{E_{fs} \times R_b \times G_t \times G_r \times \lambda^2}{16\pi^2} & if\ d \le d_0 \\ E_{mp} \times R_b \times G_t \times G_r \times h_t^2 \times h_r^2 & otherwise \end{cases} \tag{10}$$

According to [33], E_{fs}=10pJ/bit/m^2 and E_{mp}=0.0013pJ/bit/m^4 can guarantee the minimum received signal to noise ratio at the receiver required for successful information decoding (30dB). In this case, P_{r_thresh}=6.3nW (-52dBm). Correspondingly, the amount of power which can be harvested by j through the SWIPT from i is:

$$P_{EH}(j, i) = (1 - \rho) \times P_r \times \alpha \tag{11}$$

where α is the efficiency of converting the harvested energy into the stored battery energy.

For intra-cluster data transmission, only edge CMs perform intra-cluster SWIPT, which can help avoid the energy waste caused by bidirectional energy transfer. Edge CMs are defined as the CMs which are only responsible for delivering their own data to the CH but do not participate in inter-cluster data relay, and others are called nonedge CMs. According to the residual energy and expected energy consumption of the CH, its harvested energy and estimated energy consumption, each edge CM determines whether it will leverage intra-cluster SWIPT to compensate for the high energy consumption of its CH and how much the transmit power is. The expected energy consumption of a nonlayer 1 CH (say CH j) is:

$$\begin{aligned} E_{exp}(\text{CH } j) =& (E_{elec} + E_{DA}) \times L \times CMs(\text{CH } j) + (E_{elec} + E_{fs} \times d_{tonext}^2) \times L \\ &+ (2 \times E_{elec} + E_{fs} \times d_{tonext}^2) \times L \times \frac{l_m^2 - l(\text{CH } j)^2}{l_m^2 - 1} \end{aligned} \tag{12}$$

where E_{DA} is the energy consumption of data aggregation per bit; $CMs(\text{CH } j)$ is the number of CMs in cluster j; d_{tonext} is the distance to the next-hop relay; $l(\text{CH } j)$ is the layer number of CH j. The 3 items on the right side of Equation (12) denote the average energy consumption of intra-cluster data reception and aggregation, transferring its own data to the next hop and relaying packets for others, respectively. Each edge CM makes decisions according to the following rules:

① Case 1: $E_{residual}(\text{CH } j) \ge E_{exp}(\text{CH } j)$. If the summation of the total energy consumption of edge CM i in the first 4 stages and the estimated energy consumption of transmitting data to its CH is smaller than its harvested energy $E_{Harv}(i)$, CM i will deliver all the excessive portion of its harvested energy to its CH, otherwise it will only transmit data.

② Case 2: $E_{residual}(\text{CH } j) < E_{exp}(\text{CH } j)$. If the residual energy of edge CM i is higher than its estimated energy consumption, the transmit power of i $P_{send}(i)$ should satisfy

Equation (13), otherwise i will only transfer data.

$$E_{residual}(\text{CH } j) - \frac{E_{exp}(\text{CH } j)}{edgeCMs(\text{CH } j)} + P_{EH}(\text{CH } j, i) \times t = E_{residual}(i) - E_{elec} \times L - P_{send}(i) \times t \tag{13}$$

where $edgeCMs(\text{CH } j)$ is the number of edge CMs of CH j; t is the time required to transmit data from CM i. The difference between the first 2 items on the left side of Equation (13) represents the total residual energy of CH j after obtaining the energy compensation from all edge CMs except i. Adding the energy harvested from the RF signal emitted by CM i into the left side while subtracting the energy consumption of i from the right side is helpful in balancing the residual energy between CH j and CM i. $P_{send}(i)$ can be obtained by combining Equations (11)-(13), and the result is:

$$P_{send}(i) = \left(\frac{d^2_{i,\text{CH } j}}{\alpha + d^2_{i,\text{CH } j}}\right) \times \left[P_{r\_thresh} \times \alpha + \frac{E_{residual}(i) - E_{residual}(\text{CH } j) - E_{elec} \times L + \frac{E_{exp}(\text{CH } j)}{edgeCMs(\text{CH } j)}}{t}\right] \tag{14}$$

where $d_{i,\text{CH } j}$ is the Euclidean distance between CM i and CH j.

For inter-cluster data transmission, only edge CHs perform inter-cluster SWIPT, and they are defined as the CHs which only deliver their own data towards the sink but do not take part in data relay. Similar to edge CMs, the decision rules for edge CH s (its next-hop relay is q) are listed below:

① Case 1: $E_{residual}(q) \geq (2E_{elec} + E_{fs} \times d^2_{tonext}) \times L$. Here, d_{tonext} is the Euclidean distance from q to its next-hop relay. If the summation of its harvested energy and the energy obtained through intra-cluster SWIPT is higher than its energy consumption, edge CH s will deliver all the excessive energy to q, otherwise s will only transmit data.

② Case 2: $E_{residual}(q) < (2E_{elec} + E_{fs} \times d^2_{tonext}) \times L$. If the summation of the harvested energy and the energy obtained through intra-cluster SWIPT is higher than the energy consumption, the transmit power of s $P_{send}(s)$ should satisfy Equation (15), otherwise s will only deliver data.

$$E_{residual}(s) - E_{elec} \times L - P_{send}(s) \times t = E_{residual}(q) - (2 \times E_{elec} + E_{fs} \times d^2_{tonext}) \times L + P_{EH}(q, s) \times t \tag{15}$$

$P_{send}(s)$ is obtained by substituting Equation (11) into Equation (15), and the result is:

$$P_{send}(s) = \left(\frac{d^2_{s,q}}{\alpha + d^2_{s,q}}\right) \times \left[P_{r\_thresh} \times \alpha + \frac{E_{residual}(s) - E_{residual}(q) + (E_{elec} + E_{fs} \times d^2_n) \times L}{t}\right] \tag{16}$$

4. **Results and discussion.** By utilizing Matlab simulation, SMCRP is validated through performance comparison with CogLEACH [20], NSAC [21], WCM [22], DSAC [23] and RFMCRP [15] in terms of network lifetime, energy balancing capability and average number of effective data gathering nodes. The simulation parameter settings are shown in Table 1 [15, 33]. Network lifetime is defined as the round in which the first death node appears, and the network lifetime of all competing protocols is shown in Figure 2.

We can observe from Figure 2 that compared with RFMCRP, the first death node of SMCRP appears 64.07% later, i.e., in round 4881, and compared with other competing protocols, the advantages of SMCRP are more obvious. The reasons are analyzed as follows: SWIPT technology is leveraged by SMCRP to simultaneously transfer information and RF energy to the receiver to compensate for its energy consumption and balance their

Table 1. Simulation parameter settings

Parameters	Values
Network radius R	150m
Total number of CRSNs nodes N	200
Number of PUs P	10
Number of licensed channels C	10
Data packet size L	1000bit
Control packet size L_1	100bit
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
Power amplifier coefficient in multi-path loss model E_{mp}	0.0013pJ/bit/m^4
Maximum transmissionrange of CRSNs nodes d_n	50m
Transmit power of the sink P_{sink}	100W
Transmit power of PUs P_{PUs}	40W
Time duration of EH T	0.2s

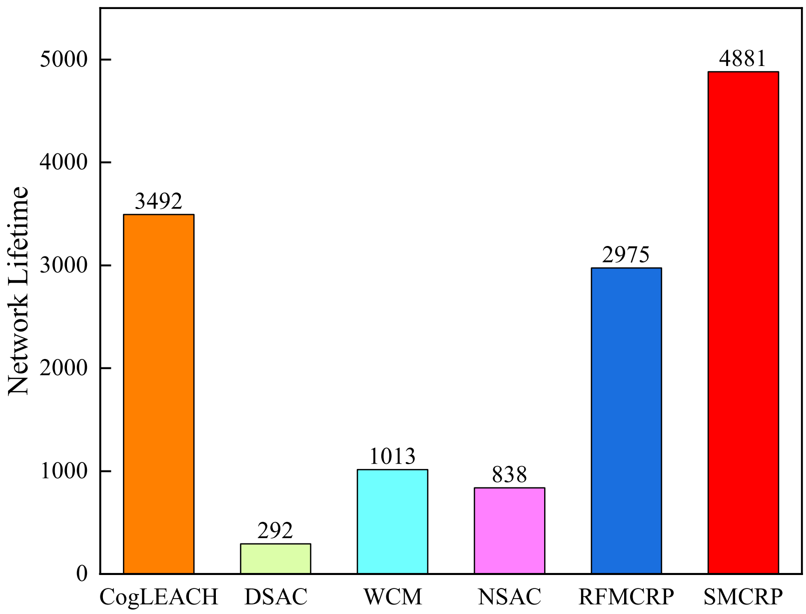

Figure 2. Comparison results of network lifetime

residual energy. Edge CMs and edge CHs will deliver their excessive portion of harvested energy to their CH or next-hop relays if they have. If the residual energy of CHs or next-hop relays is not sufficient enough to support the data transmission in this round, more energy is required from edge CMs and edge CHs to achieve intra-cluster and inter-cluster energy balance. As nonedge CMs and nonedge CHs need to consume more energy in data forwarding, in order to avoid the energy waste caused by bidirectional energy transfer, they do not need to transfer energy to their CHs or next hops. As a result, the residual energy among nodes can be well balanced, and the network lifetime can be significantly prolonged.

In order to further analyze and quantify the capability of SMCRP in balancing the residual energy among nodes, intra-cluster normalized energy balancing indicator (NEBI) and network NEBI are defined. Firstly, the residual energy of all nodes within a cluster should be normalized. For example, the normalized residual energy of CM i in cluster j can be expressed as:

$$E(i) = \frac{E_{residual}(i)}{\max\limits_{k \in clusterj} E_{residual}(k)} \tag{17}$$

where the denominator represents the maximum residual energy of all nodes in cluster j. Secondly, the average normalized residual energy of cluster j can be calculated as below:

$$aveE(j) = \frac{1}{CMs(j) + 1} \times \sum_{k \in clusterj} E(k) \tag{18}$$

where $CMs(j)+1$ is the total number of nodes in cluster j including all CMs and CH j. Thirdly, the intra-cluster NEBI in round r is defined as:

$$c - NEBI(r) = \frac{1}{|\mathbf{CHs}(r)|} \times \sum_{j \in \mathbf{CHs}(r)} \frac{\sum\limits_{k \in clusterj} (E(k) - aveE(j))^2}{CMs(j) + 1} \tag{19}$$

where $\mathbf{CHs}(r)$ is the set of CHs elected in round r. Similarly, the network NEBI in round r can be calculated according to Equation (20).

$$n - NEBI(r) = \frac{1}{|\mathbf{ClusterN}(r)|} \times \sum_{k \in \mathbf{ClusterN}(r)} (E(k) - aveClusterNE(r))^2 \tag{20}$$

where $aveClusterNE(r)$ is the average normalized residual energy of the whole network, and its calculation is similar to Equation (18) except that the residual energy of each node should be normalized by the maximum node residual energy of the whole network. $\mathbf{ClusterN}(r)$ denotes the set of CRSNs nodes which participate in clustering in round r. From Equation (19) and Equation (20), it can be seen that smaller NEBI value means stronger capability of energy balancing. The $c - NEBI$ and $n - NEBI$ of RFMCRP and SMCRP are shown in Figure 3 and Figure 4, respectively.

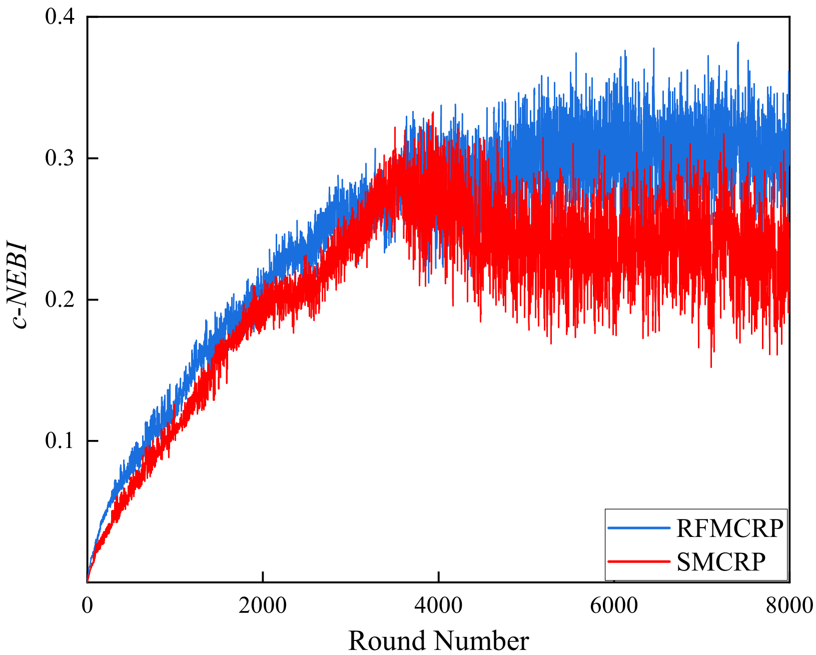

FIGURE 3. Comparison results of $c - NEBI$

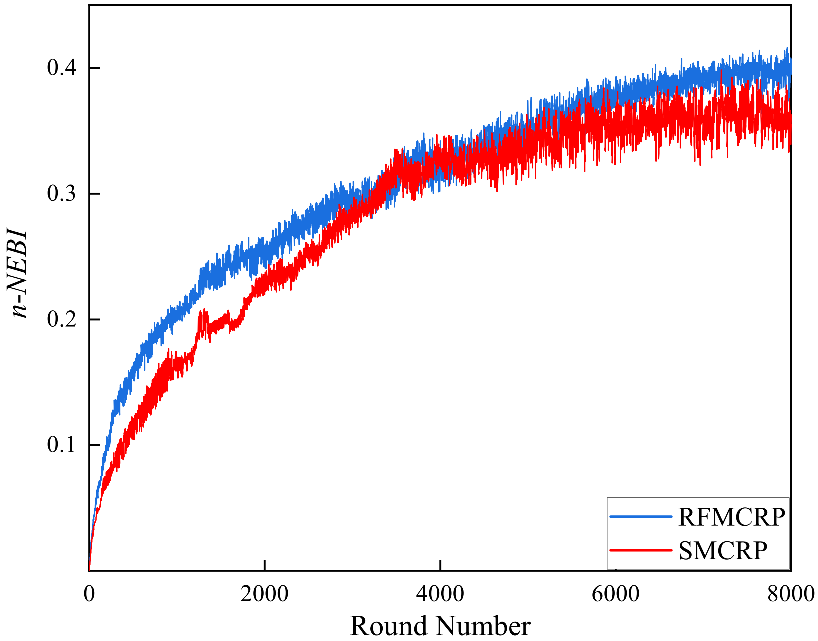

FIGURE 4. Comparison results of $n-NEBI$

From Figure 3 and Figure 4, we can see that the $c-NEBI$ and $n-NEBI$ values of SMCRP are both smaller than those of RFMCRP. To be specific, the average $c-NEBI$ of SMCRP is 0.208, which is about 15.1% lower than that of RFMCRP, and the gap becomes more obvious as network operation goes on. This results from the fact that the difference in the residual energy among nodes in the same cluster gets larger as r increases. In SMCRP, if CMs can harvest more energy to offset their energy consumption, the excessive portion of energy can be timely delivered to their CHs for compensation. If CMs are informed that the residual energy of their CHs is not sufficient enough to cover the energy cost, more energy is transferred according to the rules of energy balance. Therefore, the advantages of energy compensation are more obvious when there is less residual energy. As shown in Figure 4, the average $n-NEBI$ value of SMCRP is 0.283, which is about 7.82% smaller than that of RFMCRP. By comparison, it can be seen that $n-NEBI$ is generally higher than $c-NEBI$. The reason can be analyzed as follows: all CMs are leveraged to provide energy compensation for CHs through intra-cluster SWIPT (evaluated by $c-NEBI$) while inter-cluster energy compensation is limited to nodes on the established routes (evaluated by $n-NEBI$), which restricts the potentials of energy balancing.

Apart from the above performance metrics, the average number of effective data gathering nodes is important for evaluating the network surveillance capability of clustering protocols, and the simulation results are shown in Figure 5.

From Figure 5, it can be seen that the average number of effective data gathering nodes of SMCRP is always higher than that of RFMCRP before round 3615 due to the performance enhancement brought by energy balancing through SWIPT. In round 3615, the average number of effective data gathering nodes of SMCRP is 155, which is about 4.9% lower than that of RFMCRP. Although SMCRP exploits intra-cluster and inter-cluster SWIPT to transfer RF energy among nodes to prolong the network lifetime and balance the residual energy in the whole network, the loss during energy transfer cannot be completely avoided. Therefore, compared with RFMCRP which does not require energy

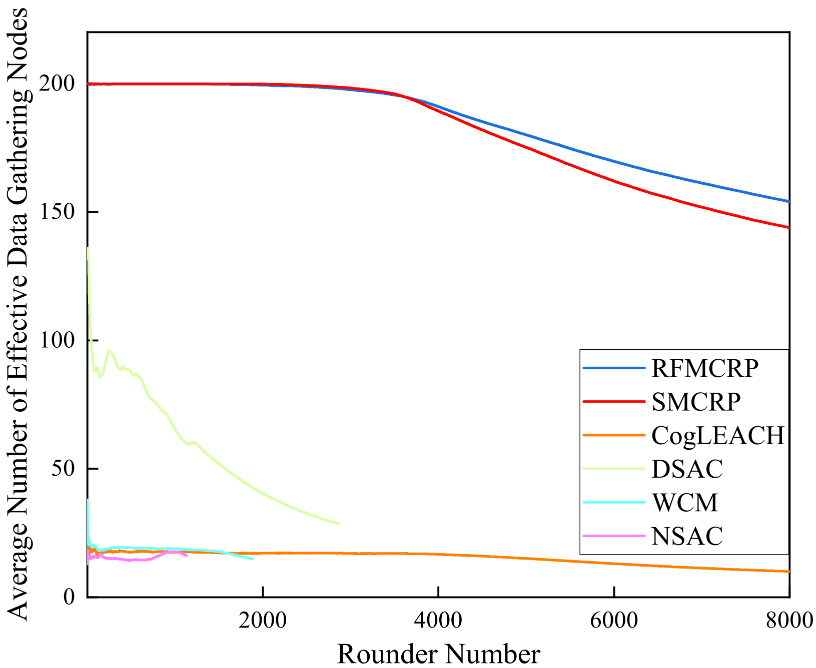

FIGURE 5. Comparison results of average number of effective data gathering nodes

transfer among nodes, there is more total energy consumption, and this is the price to pay for balancing the residual energy among nodes and prolonging the network lifetime.

5. **Conclusions.** To achieve a more balanced residual energy distribution among nodes in multi-hop RF EH-CRSNs and enable a longer network lifetime, SWIPT technology is introduced to provide energy supplement for nodes with high energy consumption, and SMCRP is proposed on the basis of our previous work. Apart from the RF energy harvested from dedicated or ambient energy sources, the energy transferred through SWIPT is included to define the criteria for CHs selection, cluster formation and multi-hop inter-cluster route selection. Simulation results demonstrate that SMCRP achieves better performance in terms of the network lifetime and energy balance than other competing protocols. To be specific, compared with RFMCRP, the network lifetime is extended by 64.07% while intra-cluster energy balance degree and network energy balance degree are improved by 15.1% and 7.82%, respectively. SMCRP is designed on the basis of perfect spectrum sensing and linear EH model. Perfect spectrum sensing assumes that there is no sensing error, i.e., there is no false alarm and missed detection. Here, false alarm represents that the spectrum is not occupied by PUs, but secondary users (SUs) mistakenly perceive that PUs are occupying it. Missed detection means that the spectrum is being occupied by PUs, but SUs have not detected it. In practical communication system, noise, sudden signal interference and feedback information error will affect the accuracy of spectrum sensing, which may result in collisions with PUs, transmission failure or constrained spectrum usage. Therefore, the impact of imperfect spectrum sensing should be considered in clustering protocol design to improve successful data delivery. In addition, the linear EH model used in this paper assumes that the output power P_{out} of the EH circuits increases linearly with the input power P_{in}. In fact, the nonlinear characteristics of end-to-end energy conversion introduced by nonlinear components in practical EH circuits such as diodes lead to the gradual saturation of P_{out} with the increase of P_{in}. At present, researchers have proposed various nonlinear EH models, such as exponential

model, power law model, saturated model and piecewise linear model, to quantify the nonlinear relationship between P_{in} and P_{out}. Therefore, in the future work, we will select the most suitable nonlinear EH model and design clustering routing protocols for RF EH-CRSNs based on it.

Acknowledgement. This work is partially supported by the National Natural Science Foundation of China under grant 61901102. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] X. 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 and Sciences*, vol. 135, no.1, pp. 711-732, 2023.

[2] 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.

[3] 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.

[4] E. Dixit, and V. Jindal, "IEESEP: an intelligent energy efficient stable election routing protocol in air pollution monitoring WSNs," *Neural Computing and Applications*, vol. 34, pp. 10989–11013, 2022.

[5] 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.

[6] K. Sinha, and B. P. Sinha, "Challenged networks to challenged computing: an untapped potential for future space exploration," in *2021 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, Cleveland, OH, USA, pp. 14-19, 2021.

[7] 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.

[8] M. Zheng, C. Q. Wang, M. Song, W. Liang, and H. B. Yu, "SACR: a stability-aware cluster-based routing protocol for cognitive radio sensor networks," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17350-17359, 2021.

[9] A. lrabea, O. A. Alzubi, and J. A. Alzubi, "A task-based model for minimizing energy consumption in WSNs," *Energy Systems*, vol. 13, pp. 671-688, 2022.

[10] S. Chaurasia, K. Kumar, and N. Kumar, "MOCRAW: A meta-heuristic optimized cluster head selection based routing algorithm for WSNs," *Ad Hoc Networks*, vol. 141, 103079, 2022.

[11] E. B. Ali, S. Kishk, and E. H. Abdelhay, "Multidimensional auction for task allocation using computation offloading in fifth generation networks," *Future Generation Computer Systems*, vol. 108, pp. 717-725, 2020.

[12] Z. Y. Shi, X. Z. Xie, H. B. Lu, H. L. Yang, J. Cai, and Z. J. Ding, "Deep reinforcement learning-based multidimensional resource management for energy harvesting cognitive NOMA communications," *IEEE Transactions on Communications*, vol. 70, no. 5, pp. 3110-3125, 2022.

[13] S. C. Zhang, Y. H. Wang, Y. W. Zhang, P. Wan, and J. W. Zhuang, "Riemannian distance-based fast k-medoids clustering algorithm for cooperative spectrum sensing," *IEEE Systems Journal*, vol. 16, no. 1, pp. 880-890, 2022.

[14] R. Lyu, W. C. Cheng, B. Z. Shen, Z. Y. Ren, and H. L. Zhang, "OMA-SWIPT for IoE-driven 6G," *IEEE Communications Magazine*, vol. 60, no. 3, pp. 19-25, 2022.

[15] J. H. Wang, and Y. Y. Ge, "A radio frequency energy harvesting-based multihop clustering routing protocol for cognitive radio sensor networks," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7142-7156, 2022.

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

[17] 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.

[18] J. H. Wang, S. Li, and Y. Y. Ge, “Ions motion optimization-based clustering routing protocol for cognitive radio sensor networks,” *IEEE Access*, vol. 8, pp. 187766–187782, 2020.

[19] J. H. Wang, and C. H. Li, “A weighted energy consumption minimization-based multi-hop uneven clustering routing protocol for cognitive radio sensor networks,” *Scientific Reports*, vol. 12, no. 1, 14039, 2022.

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

[21] 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.

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

[24] 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.

[25] G. A. Shah, F. Alagoz, E. A. Fedal, and O. B. Akan, “A 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.

[26] S. Aslam, W. Ejaz, and M. Ibnkahla, “Energy and spectral efficient cognitive radio sensor networks for Internet of Things,” *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3220-3233, 2018.

[27] 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, 107028, 2020.

[28] 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.

[29] 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.

[30] J. H. Wang, H. Ni, Y. Y. Ge, and S. Li. “Traffic-driven ions motion optimization-based clustering routing protocol for cognitive radio sensor networks,” *PLOS ONE*, vol.17, no. 9, e0272505, 2022.

[31] K. W. Choi, S. I. Hwang, A. A. Aziz, H. H. Jang, J. S. Kim, D. S. Kong, and D. I. Kim. “Simultaneous wireless information and power transfer (SWIPT) for Internet of Things: novel receiver design and experimental validation,” *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2996-3012, 2020.

[32] X. Zhou, R. Zhang, and C. K. Ho. “Wireless information and power transfer: architecture design and rate-energy tradeoff,” *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754-4767, 2013.

[33] 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.