

Article

In-Network Data Aggregation for Ad Hoc Clustered Cognitive Radio Wireless Sensor Network

Mohamad Rida Mortada ^{1,2}, Abbass Nasser ^{1,2,*} , Ali Mansour ¹  and Koffi-Clément Yao ³

- ¹ LABSTICC UMR CNRS 6285, ENSTA-Bretagne, 29806 Brest, France; mohamad.mortada@ensta-bretagne.org (M.R.M.); ali.mansour@ensta-bretagne.fr (A.M.)
- ² ICCS-Lab, Computer Science Department, American University of Culture and Education, Beirut 1507, Lebanon
- ³ LABSTICC UMR CNRS 6285, Université de Bretagne Occidentale, 29238 Brest, France; koffi-clement.yao@univ-brest.fr
- * Correspondence: abbass.nasser@ieee.org

Abstract: In cognitive radio wireless sensor networks (CRSN), the nodes act as secondary users. Therefore, they can access a channel whenever its primary user (PU) is absent. Thus, the nodes are assumed to be equipped with a spectrum sensing (SS) module to monitor the PU activity. In this manuscript, we focus on a clustered CRSN, where the cluster head (CH) performs SS, gathers the data, and sends it toward a central base station by adopting an ad hoc topology with in-network data aggregation (IDA) capability. In such networks, when the number of clusters increases, the consumed energy by the data transmission decreases, while the total consumed energy of SS increases, since more CHs need to perform SS before transmitting. The effect of IDA on CRSN performance is investigated in this manuscript. To select the best number of clusters, a study is derived aiming to extend the network lifespan, taking the SS requirements, the IDA effect, and the energy consumed by both SS and transmission into consideration. Furthermore, the collision rate between primary and secondary transmissions and the network latency are theoretically derived. Numerical results corroborate the efficiency of IDA to extend the network lifespan and minimize both the collision rate and the network latency.

Keywords: cognitive radio; wireless sensor network; in-network data aggregation; network lifespan; multihop routing



Citation: Mortada, M. R.; Nasser, A.; Mansour, A.; Yao K. C. In-Network Data Aggregation for Ad Hoc Clustered Cognitive Radio Wireless Sensor Network. *Sensors* **2021**, *21*, 6741. <https://doi.org/10.3390/s21206741>

Academic Editor: Francesca Cuomo

Received: 28 August 2021

Accepted: 2 October 2021

Published: 11 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Currently, wireless sensor networks (WSNs) play an important role in several applications generating a large amount of data. Various applications require continuous updates of critical information—for example, the video streaming of security cameras. These applications need huge spectrum resources. However, the radio spectrum becomes congested by different applications and available channels become scarce. Cognitive radio (CR) has gradually emerged in order to solve the channel scarcity problem [1–3].

CR ensures a high spectral efficiency by sharing the frequency band between two kinds of users: primary user (PU) and secondary user (SU). In CR networks, a SU can access, in an opportunistic way, an idle channel licensed to a PU. In order to avoid any harmful interference to the PU, the SU must sense the channel to keep aware of the PU status. When the PU resumes its activities, the SU should immediately vacate the channel. Therefore, the SU should be equipped with a spectrum sensing (SS) unit [4–6]. IEEE 802.22 is a CR-based standard used for CR applications. It uses white spaces in the television (VHF/UHF TV broadcast bands) within various frequency bandwidths, such as 54 MHz and 862 MHz, potentially in the 1300 to 1750 MHz and 2700 to 3700 MHz [7].

Applying CR in WSN results in the cognitive-radio-based WSN (CRSN), where the nodes behave as SUs. Thus, they should be aware of their channel status by performing

SS to monitor the PU activity. Even the incorporation of CR capability in WSN enhances the spectrum efficiency by exploiting the unused or less used spectrum; this efficiency increase comes at the cost of an additional power consumption due to the performed sensing operations [8,9].

In WSN, clustering, multihop routing (MHR), and data aggregation are efficient techniques adopted to reduce the transmission energy and to maintain the transmission quality and reliability [10–14]. They reduce the transmission radius of the nodes, as each one sends its data via the cluster head (CH) to the central base station (BS). Each node in a CRSN may perform various physical sensing, such as temperature, humidity, moisture, etc., and then transmits its data toward its designated CH. Distances among the nodes and their CHs can be decreased by increasing the number of clusters in the WSN. In addition, clustering decreases the redundancy of the nodes' data, since the CH may apply an in-network data aggregation (IDA) procedure (based on the mean, the max, or the min of the received data [15]) from the nodes and only transmits aggregated data, using MHR, to the next nearest CH toward the BS, which is the final destination. Like WSN, CRSN nodes consume energy to perform several operations: physical sensing, clustering election, data transmission, etc. In addition to these operations, CRSN consumes energy to perform SS. This inevitable function becomes an energy consuming challenge for the CRSN. Our work focuses on the two potential energy consumption sources, SS and data transmission. Other energy consumption sources are ignored, since the data transmission is the most consuming one in a traditional WSN [16]. In our model, the SS is the exclusive responsibility of CH. Before gathering the data from the cluster nodes and previous CHs, CH should perform SS to avoid any interference with the PU. As mentioned previously, a small number of clusters leads to high energy consumption during the data transmission phase due to the distance among CHs. On the other hand, a big number of clusters leads to drain in energy during the SS phase. Furthermore, the CHs perform IDA in order to reduce the data redundancy. This will impact both the size of the data and the transmission duration. Accordingly, the collision rate between the PU and SU transmissions is influenced by the data aggregation as well as the latency of the data delivery.

In this paper, we investigate the optimization of the CHs number within a CRSN given several parameters, such as the IDA, the energy consumption by physical sensing, energy consumed during the transmission, the SS performance that fulfils the predefined protection level for the PU against the interference of the CRSN nodes, and the high spectrum efficiency to be exploited by these nodes. The existing research works in the literature did not explicitly address the role of SS in determining the optimal number of clusters in a CRSN. The SS requirements should be respected since the CRSN operates as secondary network, and it should respect minimum tolerance of collision with the primary network. Furthermore, to the best of the authors' knowledge, this is the first research study that addresses the impact of IDA on the collision rate between primary and secondary transmission, the average number of the relayed packets by each CH, and the network latency. The main contributions of this paper are summarized as follows:

1. Seeking an optimal number of clusters of CRSN that takes into consideration several parameters, such as the energy efficiency related to the transmission and the sensing operations, detection, and false alarm probabilities of clustered CRSN.
2. The average number of relayed packets by each CH is obtained in a closed form.
3. Probability of collision and the latency of our energy-optimal clustered CRSN are derived in closed forms, taken into consideration the data redundancy reduction thanks to the IDA.
4. Numerical results are calculated to show the efficiency of choosing the optimal number of clusters, and to figure out the related collision rate and the latency.

The rest of the paper is organized as follows. In Section 2, we present some related works. In Section 3, we present the adopted system model with essential parameters related to the nodes' distribution, the ad hoc communication, and the data reduction of the transmitted data due to their similarity among neighbor clusters. Problem formulation

is detailed in Section 4, where the optimal number of clusters within a CRSN is derived in addition to the optimal value of the consumed energy of the node and the lifespan of the network. In Section 5, mathematical closed forms for both the collision probability and the data delivery latency are given. Numerical results are shown in Section 6 and they corroborate the efficiency of adopting the strategy of selecting the optimal number of clusters, and reveal critical parameters, such as the lifetime of the network, the collision, and the latency. In Section 7, we conclude our work and introduce future perspectives.

2. Related Works

By integrating CR features in WSN, sensor nodes in the CRSN can make the best use of available spectrum resources. However, other challenges have arisen that necessitate investigation in order to identify effective solutions. Clustering, routing, and data aggregation are all significant areas of research and solution development.

Clustering and MHR based networks are addressed in several works in WSN. In [17], the authors proposed a protocol to improve the setting up cluster and the data transmission route. Their simulations showed the effectiveness of the proposed algorithm, which extends the network lifetime and reduces the energy consumption. Thus, the authors of [18] proposed a clustering routing algorithm that prohibits nodes with a low residual energy from becoming CH. In addition, the residual energy is assumed to be an energy control factor to balance the energy consumption.

In [19], the authors propose an adaptive clustering routing algorithm for WSNs to solve the scalability problem in a large-scale case for WSNs. This algorithm divides the network nodes into clusters; then, it selects one node inside each cluster to play the role of CH. Finally, MHR paths are set according to the energy balance principle of the algorithm based on the residual energy, the load, and the number of nodes in the network. In [20], the authors use the Gustafson–Kessel algorithm, which optimizes the number of clusters in order to reduce the energy consumption based on the shape and the volume of clusters, the initial setup of a clustering algorithm, the distribution of the data objects, and the number of clusters. The authors of [21] showed that the optimal number of cluster heads (ONCH) algorithm over LEACH (the low-energy adaptive clustering hierarchy algorithm) is better than LEACH without ONCH as the energy consumption is reduced and the life of the sensor network is extended.

Clustering and MHR have also been addressed for CRSN due to their effect on the secondary interference to the PU as well as the energy efficiency. The authors of [22] proposed a novel clustering process for CRSN in which a pair of nodes close to each other and sharing the same channel can be coupled together. One of these two nodes should go into sleep mode while the other one should stay awake during a single communication interval. Then, they exchange their states after each interval. This technique improves the energy efficiency, as only the awake nodes perform the SS and participate in the election of the cluster head. The authors of [23,24] used the optimal cluster number to improve their groupwise spectrum-aware clustering algorithm. In [23], the authors proposed a novel distributed spectrum-aware clustering (DSAC) algorithm in order to make the spectrum-aware clustering more reliable and practical, while in [24], the authors suggested an optimal cluster number to build up their energy-aware cluster-based routing protocol. In these works, the optimal number of clusters was obtained based on the energy consumed in an intracluster communication related to a phase of gathering data in each cluster, and the energy consumed while transmitting data in intercluster from each CH to the BS in a single-hop transmission. Using MHR [11], the number of relays increases with the hops between sources and destinations [25].

In our work, CH plays the role of a relaying node. Therefore, in order to reduce the transmission range, we must increase the number of clusters and CHs in the network. On the one hand, the energy consumption decreases in MHR-based networks because the radio transmission distance decreases. The collision with the PU may be reduced due to restricted transmission range. On the other hand, the latency and the error probability

may increase [26]. However, another factor should be taken into consideration in CRSN, which is the SS. Thus, the optimal number of clusters in a CRSN should be obtained by optimizing the energy consumption, but also, by maintaining a target performance of the SS, the energy consumption in our system is compared to the energy consumption when using the optimal number of cluster calculated in DSAC.

On the other side, data aggregation is one of the most important compression techniques used for wireless networks. This technique reduces the energy consumption during the data gathering phase and accelerates it due to parallel data collection (each CH collects data independently and simultaneously with other CHs in the network). Data aggregation for CRSN was addressed in several works such [27–30]. The authors of [27] incorporated the basic CRSN characteristics in an information theoretical capacity maximization framework with the combination of energy-adaptive mechanisms and the information correlation for a multihop CRSN topology. A small network is used in their simulations (three sensors by hop and three hops toward the sink); it was concluded that an energy-adaptive mechanism outperforms a nonadaptive mechanism. They also concluded that the networks that use information correlation in data aggregations outperform the other networks. Works of [28–30] addressed the minimum latency data aggregation scheduling (MLDAS). In [28], a dense CRSN was considered. The authors assumed an unsymmetrical communication links among SUs and a possible SU–SU collision. To reduce the number of collisions and better sharing of the spectrum resources, SU should be in a competitor mode. Two practical distributed algorithms under a unit disk graph interference model and physical interference model were proposed in order to investigate the MLDAS problem. The authors of [29] focused on a probabilistic network model, i.e., links are not always guaranteed between nodes. Two networks are considered: a regular wireless network (RWN) with regular users and an auxiliary wireless network (AWN) with auxiliary users. Any user may send its data over the RWN spectrum with an equal chance with other users, or it can opportunistically operate on AWN if and only if the AWN spectrum is not occupied by any auxiliary user. The authors of [30] constructed the aggregation tree and the computation of a conflict-free schedule simultaneously without any predetermined structure. MLDAS was extended to multiple-channels access. Additionally, a scheduled node cannot participate in the aggregation process; thus, a new aggregation mode called data aggregation scheduling in the dark was proposed in order to utilize the spectrum opportunities of scheduled nodes.

As previously stated, numerous researchers have proposed alternative approaches to tackle major challenges facing the CRSN system. One can also notice the tremendous efforts and the persistent pursuit to reduce energy consumption and maximize the lifespan of the CRSN. In this paper, we introduce the energy consumed during the SS procedure into an optimization problem in order to find the best number of clusters that extends the network lifespan. As we will see later, SS is a costly operation that, if not properly considered, might result in an early energy depletion.

3. CRSN's System Model

Let us consider a CRSN operating opportunistically over a licensed channel reserved to a PU. For the sake of simplicity, we assume a deterministic network model, where the node links are stable, and symmetrical communication links between SUs are considered. The CRSN contains N homogeneous sensor nodes [31] distributed over a large circular geographical zone [32] with a radius R . In this zone, all physical records should be transmitted toward a BS located at the center [33]. The CRSN is assumed operating on a band allocated to a PU, i.e., the PU activity throughout all the CRSN is homogeneous. Such an assumption is suitable to the realistic scenarios means that the PU coverage is large and the whole CRSN falls into this coverage, e.g., the Global System Mobile (GSM) network [34]. Another case study is the radio FM or TV white space, where the same frequency may be used by the same PU within a whole country [35].

As shown in Figure 1, the network is divided into a set of k rings, called SS rings in reference to spectrum sensing. Rings are spaced by $\frac{R}{k}$, and each ring is divided into several

clusters called SS clusters with equal size. We suppose that the first ring (near the BS) contains three clusters; then, the area of each cluster is equal to: $\frac{\pi}{3} \left(\frac{R}{k}\right)^2$. The area of the i th ring can be evaluated as follows:

$$\begin{aligned} A_i &= \pi \left(\frac{iR}{k}\right)^2 - \pi \left(\frac{(i-1)R}{k}\right)^2 \\ &= \pi(2i-1) \left(\frac{R}{k}\right)^2 \end{aligned} \quad (1)$$

where $i \in [1, k]$ is the index of the ring. Hence, the i th ring contains a total number m_i of clusters equals to:

$$m_i = \frac{\pi \left(\frac{R}{k}\right)^2 (2i-1)}{\frac{\pi}{3} \left(\frac{R}{k}\right)^2} = 3(2i-1) \quad (2)$$

Notice that the closest ring to BS ($i = 1$) contains three clusters.

Using Equation (2), we can find the total number of clusters M in the overall network:

$$M = \sum_{i=1}^k 3(2i-1) = 3k^2 \quad (3)$$

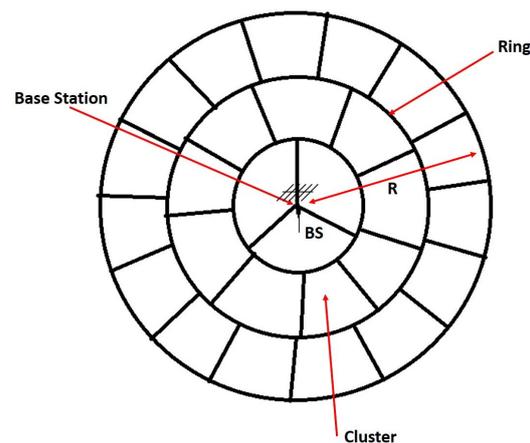


Figure 1. Clustering network with equal cluster size where $k = 3$ is the number of rings splitting a network of radius R .

At this point, the clustering algorithm with all its features (clustering mechanism, clustering cycle, and wake-up scheme) is beyond the scope of our investigation. In our research, we assume that clusters are initially existing and evenly distributed over the whole network. The topology of our adopted system model is inspired from the behavior of the well known K -means clustering algorithm when applied to a circular WSN. Figure 2 gives an example of K -mean clustering, in which the topology may be approximated to our model with three rings. As shown in this figure, the ring around the origin (where the BS is located) contains three clusters, similarly to our system model.

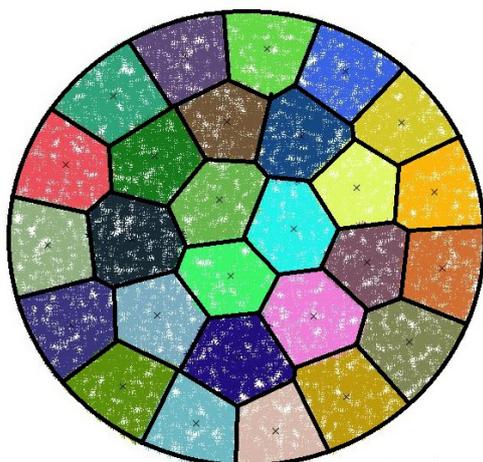


Figure 2. *K*-means applied on a circular CRSN uniformly distributed; the number of cluster was predefined.

Our network is considered as a homogeneous network. However, we can also extend our work on a heterogeneous network with A nodes where only N nodes, $N \leq A$, can be elected as CHs, because they have the required capability to fulfill the CH role.

In our model, a multihop transmission system is adopted among the CHs. Thus, the message sent from a far CH to BS should pass by the CHs in between.

According to our model, CRSN carries out an IDA technique [36], where CH gathers the data of the nodes inside its cluster and aggregates the data collected from farther CHs. Figure 3 shows an example of the process of collecting data. Nodes, other than CHs, are called cluster nodes (CNs).

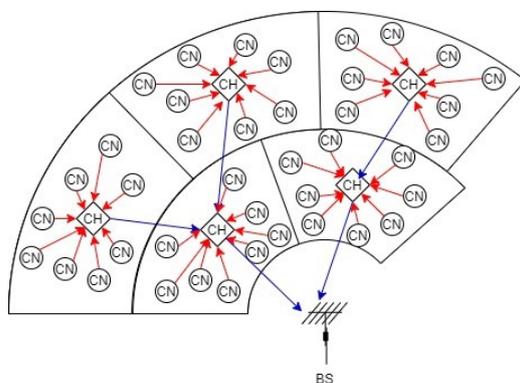


Figure 3. Gathering data from CNs and CHs.

CNs collect data on a regular basis and send it to CHs whenever requested. In turn, CH makes SS, if the channel is free; then, it requests data from their CNs as well as from neighboring CHs. All CNs transmit their packets toward their appropriate CH. In order to avoid collision among CHs transmission, CSMA/CA and TDMA techniques can be used [37].

Generally, CNs and CHs close to each other may collect similar physical values, such as humidity, temperature, etc. Figure 4 shows a part of CRSN divided into clusters and virtual physical zones with similar clusters (dashed lines) that have similar data packets. The network is also divided into virtual physical rings, and contains k_p physical rings. Each physical zones contains $\frac{k}{k_p}$ rings where $k_p \leq k$. In average, each physical ring contains a number of physical zones, formed by an equal number of SS clusters. Let our CRSN has P similar physical value zones generating in average P distinct packets in the entire CRSN. Each physical zone contains several SS clusters. Hence, the maximum number of distinct

packets retransmitted by the nearest CHs to the BS is $\frac{P}{3}$, where 3 is the number of clusters corresponding to the first ring (see Figure 1). In our system, a cluster receives the packets coming from its previous CHs' peers. Then, it compares received packets with each other and with its own packet collected from its cluster members. Subsequently, it compacts the similar packets into a single packet by counting the number of the occurrence and merging them into a new packet before sending it. In the end result, this cluster sends packets that differ from each other by a predetermined threshold amount. There are many methods for calculating the difference between data packets; the method used in our previous work based on the Euclidean distance can be used such as [38]. Figure 4 shows that if a CH receives p data packets from a neighbor CH located in the same physical zone, then it retransmits p packets to the next CH, whereas if a CH receives p data packets from a CH in a different physical zone, it then retransmits $p + 1$ data packets to the forward CH [39].

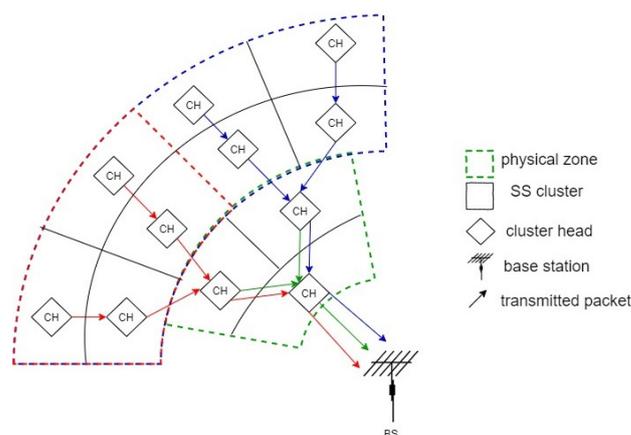


Figure 4. Data routing between CHs located in different similar physical value zones.

To detail the IDA process above, after collecting data coming from its CNs and neighboring CHs, a CH performs three steps to reduce the data redundancy, as described in our previous study [38]:

1. CH aggregates the data received from CNs in a fixed size packet [38].
2. The CH applies a similarity test performing IDA process, based on the Euclidean's norm, applied to all packets.
3. Finally, a CH just transmits distinct packets toward the next CH.

Obviously, CHs are prone to depletion of energy first. Therefore, the CH role is periodically rotated among potential nodes [12,32]. The number of hops from the CH to the BS depends on the distance between them. The ring number in the CRSN can be considered as the maximum number of hops between the farthest CH and BS. A high number of hops means that a low amount of energy will be consumed by the transmission, but the energy consumed by SS will be increased, and vice versa.

Hereinafter, we discuss how to find the optimal number of clusters that saves energy and improves the network lifespan while satisfying an acceptable level of protection for the PU against the interference characterized by a high probability of PU signal detection and achieving high spectrum efficiency by reducing the probability of false alarm on the PU presence.

4. Problem Formulation

In WSN, each node collects data from its environment and transmits them toward CH. In our model, a CH has two major roles:

1. Performing the SS before requesting the data from the CNs and its previous CHs.
2. IDA based data gathering and retransmitting from its CNs and its backward CHs as per ad hoc topology.

Therefore, the main energy consumption of the CH comes from two operations:

- When performing the SS operation, an energy amount e_s is needed.
- The transmission of the data from a given CH toward the next CH in the direction of destination also consumes energy.

The energy of data transmission depends on the decision outcome of SS performed by CH giving the permission to transmit the data:

1. CH makes a miss detection decision on the PU status, i.e., PU is detected absent while it is truly active. Thus, a collision may take place.
2. CH makes a true decision on the absence of PU. In this case, CH can send its data toward the next CH freely on the PU channel.

Accordingly, the probability that a CH transmits the data is given by:

$$p_t = p(H_0)(1 - p_{fa}) + p(H_1)(1 - p_d) \quad (4)$$

where $p(H_0)$ and $p(H_1)$ are the prior probabilities that the channel is idle or active, respectively. p_{fa} and p_d are the desired false alarm and detection probabilities, respectively, ensuring the SS requirements.

The number of SS operations and the transmission range are directly related to the number of rings k . Therefore, finding the optimal number of rings k , which extends the network lifespan L , becomes essential for the network. Furthermore, network lifespan is directly affected by the energy consumption along CHs and CNs in our network. Because of the obvious difference in power consumption between the CHs and the CNs, our problem can be formulated as finding the optimal k according to energy consumed by the CHs only.

Let us define e_i , the average energy consumed by a network node (CH or cluster member) located in the i th ring:

$$e_i = \left[e_r \left(\frac{R}{k} \right)^2 p_t N_{p_i} + e_s + e_0 N_{p_i} \right] \frac{M}{N} + \left[e_r \left(\frac{R}{k} \right)^2 p_t + e_0 \right] \left(1 - \frac{M}{N} \right) \quad (5)$$

where N_{p_i} is the average number of packets relayed by CHs in the i th ring, e_r is the energy density at the receiver antenna and $e_r \left(\frac{R}{k} \right)^2$ is the amount of energy consumed by CH during the data transmission. N is the total number of nodes and $\frac{M}{N}$ denotes the probability of a node to be a CH, and the term multiplied by $\frac{M}{N}$ is the energy consumed by a CH (transmission, SS, running time), whereas $\left(1 - \frac{M}{N} \right)$ is the probability of a node to be a CN, and it is multiplied by the energy consumed by the CN (transmitting data to CH, running time). e_0 is the energy consumed by the node in the running mode.

Let us define e_{max} as the maximum consumed energy per CHs in the CRSN network. In this work, the optimization is performed with respect to the first node dies (FND) criterion. Lifespan is assumed to be the number of iterations that can be performed by the CRSN before the extinction of the first CH in the network [40].

Assuming that each node is initially provided with an amount of energy called the initial energy e_{int} . Hence, lifespan L can be calculated as follows:

$$L = \frac{e_{int}}{e_{max}} \quad (6)$$

where:

$$e_{max} = \max_{1 \leq i \leq k} e_i \quad (7)$$

The optimization problem, looking for extending the network lifespan, can be defined as follows:

$$\begin{aligned} & \max_{k \neq 0} \frac{e_{int}}{e_{max}} \\ & \text{s.t. } M \leq N \end{aligned} \quad (8)$$

Since e_{int} is constant, the maximization of (8) can be reformulated as a minimization problem as follows:

$$\begin{aligned} & \min_{k \neq 0} e_{max} \\ & \text{s.t. } M \leq N \end{aligned} \quad (9)$$

Using Equation (5), we substitute e_{max} by its value in Equation (9) to obtain:

$$\begin{aligned} & \min_{k \neq 0} \left[\max_i \left(e_r \left(\frac{R}{k} \right)^2 p_t N_{p_i} + e_s + e_0 N_{p_i} \right) \frac{M}{N} + \left(e_r \left(\frac{R}{k} \right)^2 p_t + e_0 \right) \left(1 - \frac{M}{N} \right) \right] \\ & \text{s.t. } M \leq N \end{aligned} \quad (10)$$

It becomes obvious that the closest CHs to the BS (ring 1) consume the largest amount of energy due to the relaying role. Hence, our problem can be reformulated as:

$$\begin{aligned} & \min_{k \neq 0} \left[\left(e_r \left(\frac{R}{k} \right)^2 p_t k_p^2 + e_s + e_0 k_p^2 \right) \frac{M}{N} + \left(e_r \left(\frac{R}{k} \right)^2 p_t + e_0 \right) \left(1 - \frac{M}{N} \right) \right] \\ & \text{s.t. } M \leq N \end{aligned} \quad (11)$$

Proof. See Appendix C. \square

Now, replacing M , the number of CHs in the network, with its value $M = 3k^2$ as seen in Section 3 in our problem:

$$\begin{aligned} & \min_{k \neq 0} \left[\left(e_r \left(\frac{R}{k} \right)^2 p_t k_p^2 + e_s + e_0 k_p^2 \right) \frac{3k^2}{N} + \left(e_r \left(\frac{R}{k} \right)^2 p_t + e_0 \right) \left(1 - \frac{3k^2}{N} \right) \right] \\ & \text{s.t. } M \leq N \end{aligned} \quad (12)$$

Theorem 1. If $f(x)$ is a unimodal function defined on $\mathcal{G} \subset \mathbb{R}^+$ having a minimum as extremum at $x_0 > 0$, then $f(n)$, $n \in \{\mathbb{N} \cap \mathcal{G}\}$, is a set of points having a minimum $M = \min\{f(\lfloor x_0 \rfloor), f(\lceil x_0 \rceil)\}$.

where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ stands for the floor and ceiling operator, respectively.

Proof. Let $f(x)$ be a unimodal function defined on $\mathcal{G} \subset \mathbb{R}^+$ with $\frac{df}{dx}|_{x=x_0} = 0$ and $f(x_0) = \min\{f(x)\}$ for $x \in \mathcal{G}$. Let $n \in \{\mathbb{N} \cap \mathcal{G}\}$ be a natural variable and $U_n = \{f(n) \mid n \leq x_0\}$ and $V_n = \{f(n) \mid n \geq x_0\}$ are two numerical sequences, knowing that U_n is monotonically descendant, and V_n is monotonically ascendant; thus, $f(\lfloor x_0 \rfloor)$ and $f(\lceil x_0 \rceil)$ are the minimum of U_n and V_n at $\lfloor x_0 \rfloor$ and $\lceil x_0 \rceil$, respectively. Obviously, the min of $f(n)$ is $M = \min\{\min\{U_n\}, \min\{V_n\}\}$; hence, $M = \min\{f(\lfloor x_0 \rfloor), f(\lceil x_0 \rceil)\}$. \square

Using Theorem 1, we can show that our Equation (12) can be associated to a unimodal function with a single minimum on its defined domain. Thus, our problem can be simplified into a simple optimization problem. Replacing k with a continuous variable x in Equation (12), and differentiating e_i with respect to x , we obtain the optimal value:

$$k_{opt} = \operatorname{argmin} \left(e \left\lceil R^4 \sqrt{\frac{e_r p_t D}{e_s + e_0(k_p^2 - 1)}} \right\rceil, e \left\lfloor R^4 \sqrt{\frac{e_r p_t D}{e_s + e_0(k_p^2 - 1)}} \right\rfloor \right) \quad (13)$$

where $\lceil \cdot \rceil$ stands for the ceiling operator and $\lfloor \cdot \rfloor$ stands for the floor operator. if $k_{opt} \gg 1$, we can approximate k_{opt} to:

$$k_{opt} = \left\lfloor R^4 \sqrt{\frac{e_r p_t D}{e_s + e_0(k_p^2 - 1)}} \right\rfloor \quad (14)$$

where $D = \frac{N}{\pi R^2}$ is the density of the network. Since FND is adopted in this work, the optimal value e_{opt} of e_{max} is given by substituting k_{opt} in Equation (5):

$$e_{opt} = 2\sqrt{\frac{e_r p_t e_s}{D}} + (k_p^2 - 1) \left(e_0 \sqrt{\frac{e_r p_t}{e_s D}} + \frac{e_r p_t}{D} \right) \quad (15)$$

Figure 5 shows the lifetime of the nearest node with respect to the number of clusters. As it can be seen, the theoretical results coincide with the simulated ones approving Equation (5). Lifespan increases until reaching a maximum before decreasing again. Indeed, when the number of clusters is low, the size of a cluster becomes large, then nodes consume a great amount of transmission energy. Subsequently, the energy consumed while transmission becomes greater than the energy consumed to perform the SS by CH. As the number of clusters increases, their size decreases; hence, the transmission energy consumed by CHs decreases. Therefore, the lifespan increases until a maximum value at a specific value of k is reached. However, if the number of cluster continues to decrease, the energy consumed by SS is considerably high compared to the low transmission energy, so the lifespan decreases.

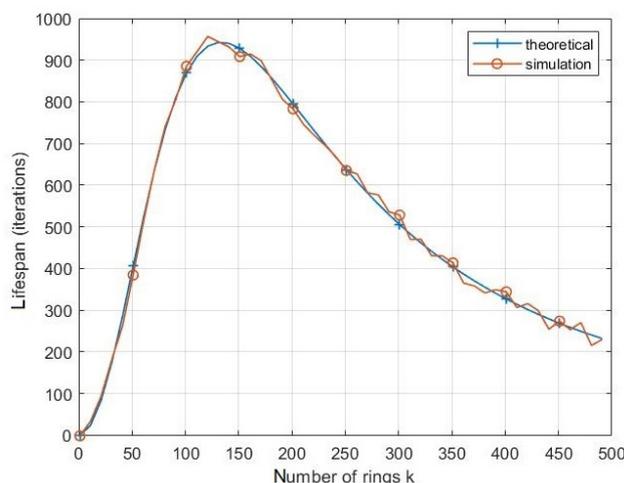


Figure 5. Theoretical and simulation results of the lifespan vs. number of rings in CRSN.

The number of clusters is related directly to the number of rings k as shown in Equation (3). Hence, the optimal value k_{opt} of k should maximize the lifespan of the network.

5. Network Latency and Collision Probability

In this section, we derive the average time taken by each round of data gathering in the network, and the probability of collision with the PU. Furthermore, using the analytical expressions of the collision and the latency, we prove analytically that the first ring ($k = 1$) of the network spends the larger amount of energy across the network.

5.1. Collision Probability and Latency

Each time the BS requests data from nodes, the CNs start sending to the CHs toward the BS. In turn, before receiving data from CNs and backward CHs, CH should sense the channel in order to ensure that PU is absent to avoid any collision.

The state of PU (active or idle) is modeled using a discrete Markov chain as shown in Figure 6 [41], where a_{ij} is the transition probability of the PU from the state i to the state j . The time in this Markov chain is divided into equal slots of τ_s seconds.

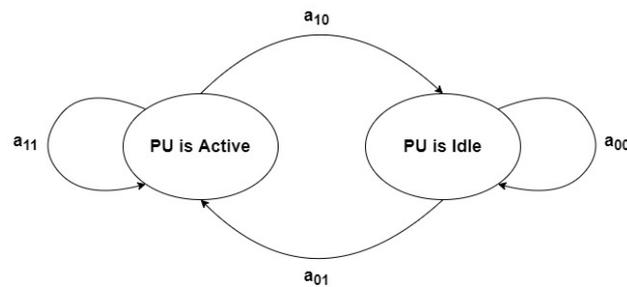


Figure 6. The PU activity modeled as a two-state Markov chain.

Each CH, acting as SU, has four tasks to do:

1. Performing SS to avoid any collision with PU.
2. Broadcasting a message to CNs and neighboring CHs indicating that it can receive collected data.
3. Gathering data from CNs and neighboring CHs and aggregating that data.
4. Waiting a broadcasting message from the next CH before transmitting packets toward it.

The data sending starts from the CHs of the farthest ring to the next one toward the BS. The probability that a collision occurs during the data transmission from the farthest ring until the BS is given by:

$$P_c = 1 - \prod_{i=1}^k (1 - P_{c_i}) \quad (16)$$

where P_{c_i} is the probability of collision with PU in the i th ring in a network. Such a collision may take place in one of the following scenarios:

1. A miss detection decision on the PU presence is made by the CH.
2. The CH makes a correct decision on the PU absence, but the latter resumes its activity during the transmission period of the CH.

Consequently, P_{c_i} could be evaluated as follows:

$$P_{c_i} = P(H_1)(1 - p_d^{m_i}) + P(H_0) \left(1 - p_{fa}^{m_i}\right) \left(1 - a_{00}^{\frac{T_{a_i}}{\tau_s}}\right) \quad (17)$$

a_{00} is the transition probability that the PU will stay in the idle state at the next time slot τ_s , and T_{a_i} is the average time of transmission activity of CH in the i th ring:

$$T_{a_i} = \frac{N \times T_{CN}}{3k^2} + \tau_{T_i} V_i \quad (18)$$

where N is the total number of CNs in the overall CRSN, T_{CN} is the period of transmission data packet from a single CN to its appropriate CH, and τ_{T_i} is the average time of collecting data of a CH in ring i from one of its V_i previous CHs in the previous SS ring $i + 1$. Equation (18) reflects the adoption of the time division multiple access (TDMA) technique by the CNs, as only one channel is assumed to be used by these nodes to send their data to CH. TDMA, in our case, prevents the interference between the transmissions of the CNs and lets the CH collect the data properly.

As mentioned previously, V_i is the average number of CHs that transmit data packets to a single CH in the ring i , in other words, the average number of previous CHs of any CH in the i th ring in our CRSN. This value can be calculated as follows:

$$V_i = \frac{3(2(i+1) - 1)}{3(2i - 1)} = \frac{2i + 1}{2i - 1} \quad (19)$$

where $i \in [1, k - 1]$.

Neighbor CNs may collect similar data. Therefore, we assume that there are P distinct physical zones with similar areas. Thus, the average time of transmission τ_{T_i} of each CH member of ring i can be presented as follows:

$$\tau_{T_i} = \frac{T_P}{m_i} \sum_{j=\lceil \frac{k_p}{k} i \rceil}^{k_p} \sum_{t=1}^{m_i} \left(\left\lceil \frac{mp_j t}{m_i} \right\rceil - \left\lfloor \frac{mp_j (t-1)}{m_i} \right\rfloor \right) \quad (20)$$

where T_P is the period of a single CH packet transmission, k_p is the number of physical ring, m_{p_j} is the number of physical zones in a physical ring j , and m_i is the number of SS clusters in the i th ring.

Proof. See Appendix A. \square

Equation (20) can be rewritten as:

$$\tau_{T_i} = \frac{T_P}{m_i} \sum_{j=\lceil \frac{k_p}{k} i \rceil}^{k_p} (m_{p_j} + m_i - m_{p_j} \wedge m_i) \quad (21)$$

with:

$$m_{p_j} + m_i - m_{p_j} \wedge m_i = \sum_{t=1}^{m_i} \left(\left\lceil \frac{mp_j t}{m_i} \right\rceil - \left\lfloor \frac{mp_j (t-1)}{m_i} \right\rfloor \right) \quad (22)$$

where $(m \wedge n)$ stands for the greatest common divisor operator of m and n .

Proof. See Appendix B. \square

Consequently, the period of overall collecting data cycle from the network becomes:

$$\begin{aligned} T_c &= \frac{1}{P(H_0)(1 - p_{fa})} \sum_{i=1}^k T_{a_i} \\ &= \frac{1}{P(H_0)(1 - p_{fa})} \sum_{i=1}^k \left[\tau_B + \frac{NT_{CN}}{3k^2} + \frac{T_P V_i}{m_i} \sum_{j=\lceil \frac{k_p}{k} i \rceil}^{k_p} m_{p_j} + m_i - (m_{p_j} \wedge m_i) \right] \quad (23) \end{aligned}$$

In Equation (23), we assume that any transmission between two CHs, or CNs and CHs, cannot succeed in presence of PU; Therefore, we divide the term $\sum_{i=1}^k T_{a_i}$ (the total time of collecting data without interruption of PU) over the probability of transmission data without presence of PU.

5.2. Maximal Consumed Energy

Regarding Equation (15) on e_{opt} , in order to prove that the CHs in the first ring consume the maximum amount of energy among all other CHs, we can instead prove that

the CHs in that ring relay the maximum number of packets toward the BS. The average number of packets relaying by CHs in the i th ring, N_{p_i} can be derived from Equation (21):

$$N_{p_i} = \frac{1}{m_i} \sum_{j=\lceil \frac{k_p}{k} i \rceil}^{k_p} (m_{p_j} + m_i - m_{p_j} \wedge m_i) \quad (24)$$

Therefore, each CH in the ring i relay N_{p_i} packets, in this case, a CH at the first ring should aggregate N_{p_1} packets:

$$N_{p_1} = \frac{1}{3} \left(\sum_{j=1}^{k_p} 3(2j-1) + 3 \sum_{j=1}^{k_p} 1 - \sum_{j=1}^{k_p} (3(2j-1) \wedge 3) \right) = k_p^2 \quad (25)$$

Hereinafter, we show that:

$$N_{p_1} > N_{p_i}, \forall i > 1. \quad (26)$$

Proof. See Appendix C. \square

6. Numerical Results

In this section, we investigate our theoretical results and show the effectiveness of the IDA in reducing both the collision probability and the network latency. The results are compared with the classical CR without IDA techniques, and we compare our energy consumption, data latency, and collision probability with the DSAC system [23].

In our simulations, we consider a circular network containing N nodes with a radius $R = 100$ m and 3 physical similar rings. According to Equation (A2) in Appendix A, the number of physical zones is $P = 27$. The processing energy of nodes during the receiving is $e_0 = 50$ nJ/packet and the transmission energy constant is $e_r = 0.1$ nJ/bit/m² [42]. For the sensing process, a SNR of -1 dB is considered, and $p_d = 0.9$ and $p_{fa} = 0.1$ are set as the target probabilities to meet the SS requirement. An energy detection method is considered for SS, and the consumed energy by the sensing e_s is evaluated according to the number of samples N_s needed for the sensing process [43] and the energy consumed per sample E_s :

$$e_s = N_s E_s \quad (27)$$

Figure 7 depicts the effects of the ring number k on the energy consumption by iteration with different values of node density D . It can be shown that increasing the density of nodes inside the network leads to decrease the energy consumed during an iteration. As for the impact of the number of rings k , the energy consumption per iteration decreases with the increase of k until the latter reaches an optimal value, at which point this energy consumption becomes minimal. If k continues increasing, the energy consumption per iteration reincreases, leading to shortening of the network lifespan.

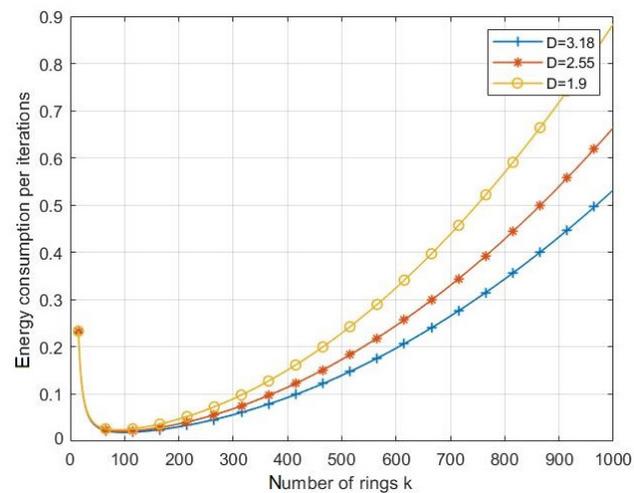


Figure 7. The effects of the ring number on the energy consumption per iteration on e_{max} with different network densities.

Figures 8 and 9 highlight the effects of the nodes number on the energy consumption and lifespan respectively for various values of k including the optimal value. As the two figures show, for all the considered values of k , the energy consumption per iteration decreases with respect to the number of nodes while the network lifespan increases. However, when setting k to its optimal value for each nodes number N , Figure 8 shows the energy consumption per iteration decreases compared to the other considered values of k , which leads to extend the network lifespan as presented in Figure 9. Note that, with a node number less than 7000, the lifespan with $k = 60$ is longer than the one observed with $k = 80$, but it appears shorter when the number of nodes is more than 7000. In fact, with less than 7000 nodes, the optimal number of k becomes closer to 60. However, this number increases with the number of nodes, and it approaches to 80 when the number of node is above 7000, as shown in Figure 10.

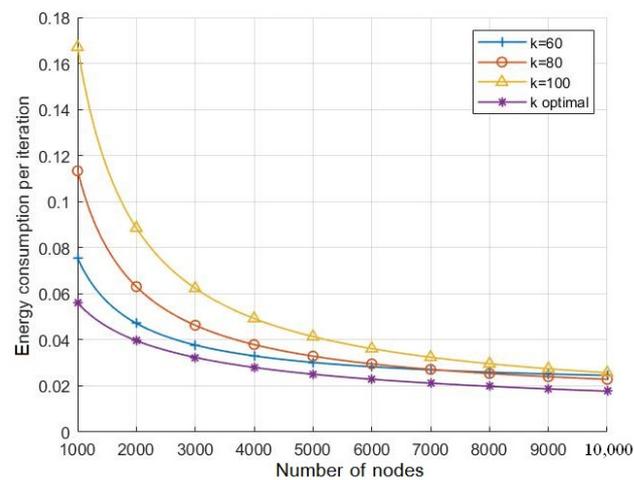


Figure 8. The effects of the number of nodes on the energy consumption with different values of k .

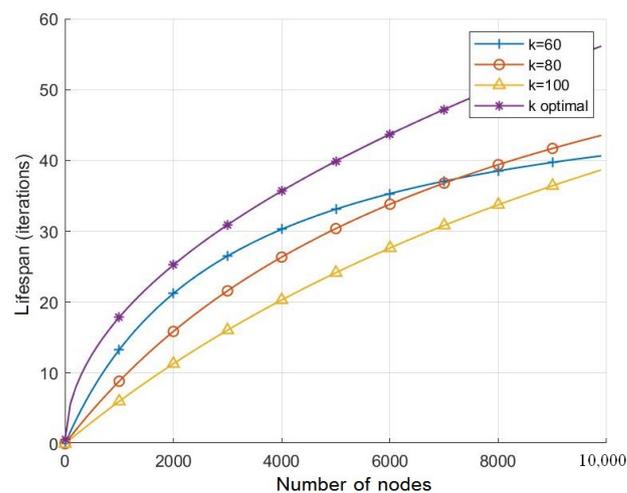


Figure 9. The effects of the number of nodes on lifespan with respect to various values of k .

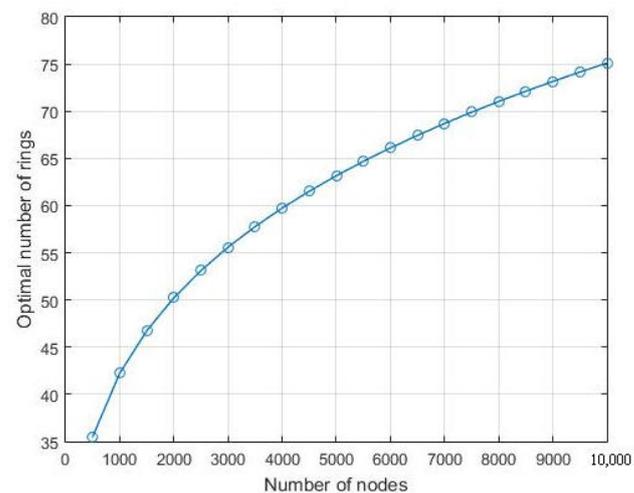


Figure 10. The optimal number of rings k vs. the node numbers.

Hereinafter, we investigate the performance of the CRSN with and without IDA in terms of the collision probability and the data delivery latency. The numerical results are based on the data size effect. Indeed, other message contents, such as the header and the trailer of the message, may affect the performance, but these contents are not considered in our work.

The collision probability of CHs with PU in the middle ring (ring $\lceil \frac{k}{2} \rceil$) of the network during the PU activity, with and without IDA, are shown in Figure 11. For both cases (with and without IDA), the collision rate increases with N due to the increasing number of messages to be transmitted. When IDA is applied, the probability of collision with PU decreases with respect to the number of rings. This effect is due to both the aggregation of the messages and the reduction of the number of CNs in each cluster. Hence, the time of collecting data in each cluster decreases leading to decrease the probability of collision. By contrast, without IDA, the probability of collision increases after reaching a minimum value. This refers to the huge number of packets retransmitted by each CH coming from the previous CHs. Without IDA, the packet length increases linearly with the number of rings and results in long transmission time of CHs.

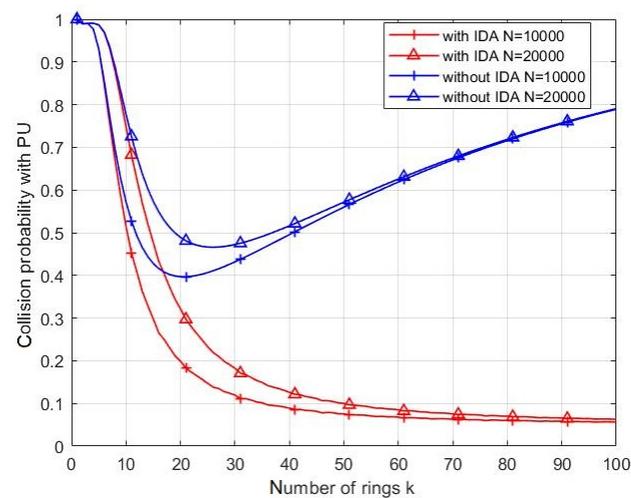


Figure 11. The collision probability of CH with PU during its activity with and without IDA.

Figure 12 shows the duration of the complete collecting data cycle from CN to the BS. As clearly shown in this figure, IDA leads the CRSN to highly reduce the data collection cycle period compared to the case without IDA. For all cases and for a small number of rings, the CH spends a great amount of time collecting data from its cluster nodes. This time decreases with an increasing k until the latter surpasses a certain value. Beyond this value of k , collecting time increases again. This effect is due to the multihop transmission, where a big number of routes results from the high values of k and procures a big time of transmission between CHs and BS.

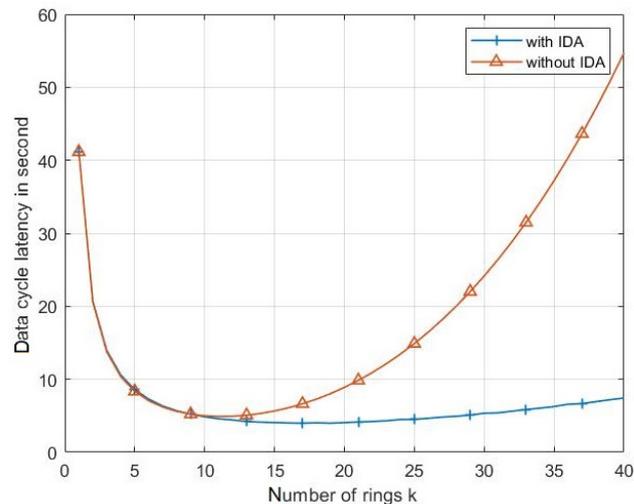


Figure 12. The duration of the complete data collection cycle from CN to the BS with and without IDA.

Figure 13 depicts the collision probability with and without IDA, during a transition of rings: $k = 1$, $k = 10$, and $k = 20$, with respect to the number of nodes in the network. Clear effect of the IDA can be noticed on reducing the collision compared to case of no IDA for the three considered values of k . For instance, for $k = 1$ and $N = 1000$, the collision rate is about 0.5 when IDA is adopted, whereas it is closed to 1 for no IDA. However, the collision rate exhibits an increasing with N for both cases due to the increase in the number of messages to send.

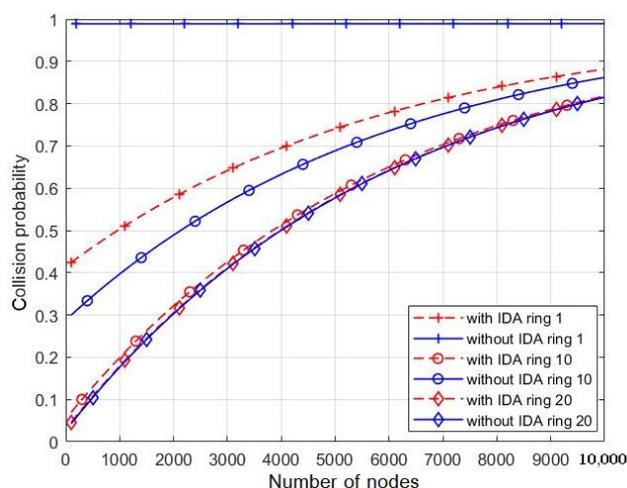


Figure 13. The collision probability of CH with PU during its activity with and without IDA with different rings.

Figure 14 shows the collision probability in networks, with and without IDA, in different rings where the number of nodes is fixed to $N = 1000$. It can be shown that, at the nearest ring from the BS (ring 1), we could notice a big difference between the two curves due to the redundancy reduction of the message obtained when IDA is applied. This must reduce the message length and, thus, alleviate the collision rate. Moving forward to the farthest ring from the BS (in our example it is ring 20), the curves begin to converge until they intersect at the last ring. This is because the collision with PU at the last ring during the SU activity is the same for both cases, with and without IDA, since there are no packets coming to the last ring, i.e., the CHs at this ring do not relay any packet from previous rings.

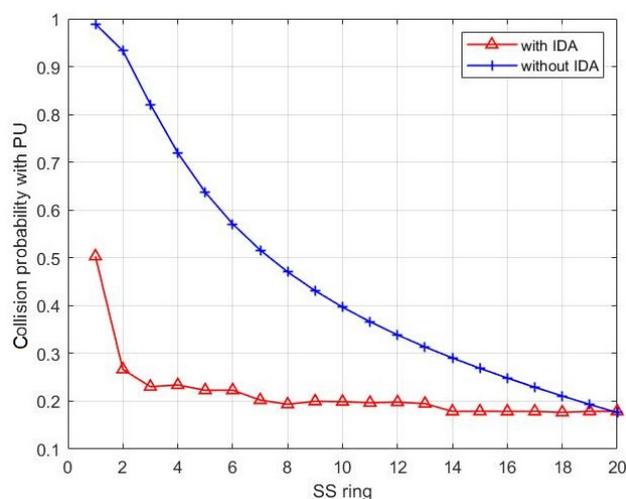


Figure 14. The collision probability of CH with PU during its activity among all network rings.

In Figure 15, we compare our system with the DSAC system mentioned in [23]. The DSAC authors have calculated the optimum value for the number of clusters in the network based on the maximum transmission distance at the sensor nodes and the density of the sensor nodes in the network. We should highlight that our simulations showed that the energy consumption in our system is significantly lower than that of DSAC. The reason here is because we have taken into consideration the amount of energy consumed during the SS process when we optimize our problem.

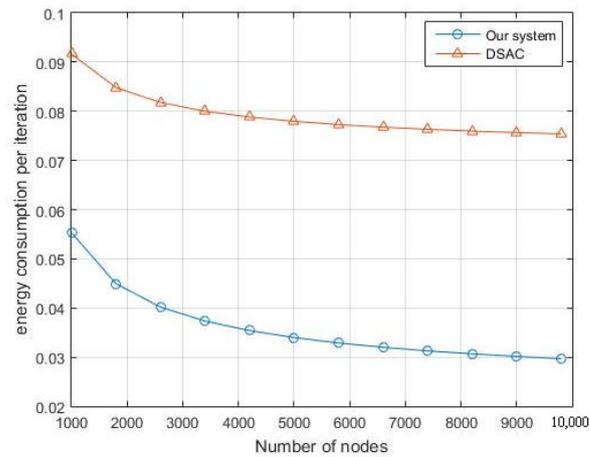


Figure 15. Comparison of energy consumption between our system and the DSAC.

We also compare the latency (Figure 16) and the collision probability with the PU (Figure 17) of the overall network in our system with the DSAC. The simulation shows that our system outperforms the DSAC in terms of collision probability and the delay.

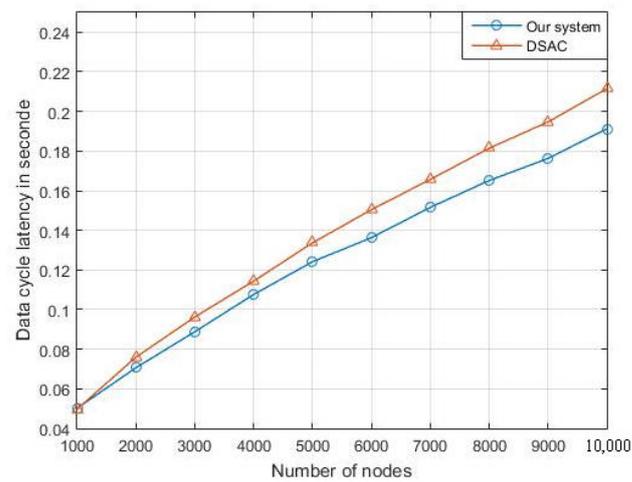


Figure 16. Comparison of data latency between our system and the DSAC.

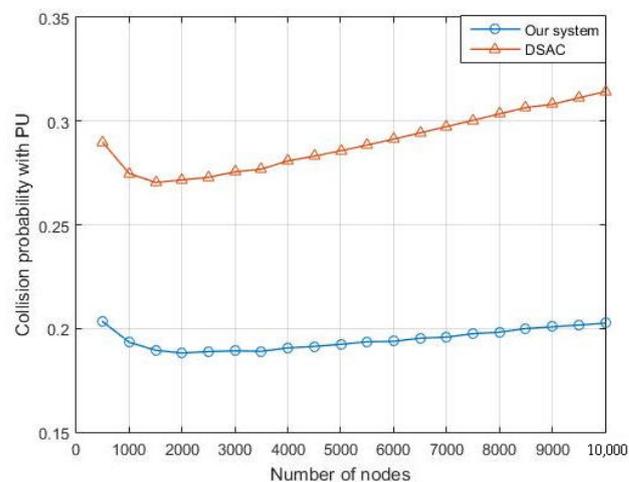


Figure 17. Comparison of collision with the PU transmission between our system and the DSAC.

7. Conclusions

In our study, we consider a CRSN where all nodes are provided with SS module. This network is divided into clusters and each cluster contains a number of nodes and one CH. The communication among CHs and the base station is multihop-based using a licensed channel dedicated to PU. SS is the responsibility of CHs. Upon a decision that PU is absent, CH broadcasts a request to its cluster's nodes and its backward CHs to gather the data. In this network, a low number of clusters leads to an increase in the energy consumption due to the data transmission, while a high number of clusters increases the energy consumption due to the SS. Thus, we calculate the optimal number of clusters that saves the consumed energy within CRSN and extends the network lifespan. Latency in delivering the message and collision rate between the CH and PU transmission are also derived given the IDA adoption by the CRSN, after calculating the average number of packets retransmitted by each CH. Numerical results show that the selection of an optimal value of the cluster number and adopting IDA may considerably extend the network lifespan.

Our work makes no attempt to quantify the energy consumed during the clustering process, although clustering may consume a significant amount of energy. Therefore, as future work, we will consider all aspects related to the clustering algorithm that will be optimized in energy consumption and simple to be implemented in CRSN. We will also derive the optimal number of clusters considering energy-harvesting-based CRSN. Moreover, it would be interesting to study the impact of number of clusters on the sustainability of the network. Another interesting perspective is to study the heterogeneous PU activity, i.e., more than one PU existing within the CRSN, on the system performance, in asymmetric links within a probabilistic network model.

Author Contributions: Conceptualization, M.R.M., A.N., A.M. and K.-C.Y.; methodology, M.R.M., A.N., A.M. and K.-C.Y.; writing—original draft preparation, M.R.M., A.N.; writing—review and editing, A.M. and K.-C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

N	Total node numbers in the network
M	Number of CHs in the network
R	Geographical zone radius
k	Number of rings in network
A_i	Area of ring i
m_i	Number of cluster inside ring i
M	Total number of clusters in the network
k_p	Number of physical rings in the network
p	Number of distinct packet received by a CH
$p(H_0)$	Probability of the PU to be idle
$p(H_1)$	Probability of the PU to be active
p_{fa}	Probability of false alarm
p_d	Probability of miss detection
p_t	Probability of transmission of a CH
N_{p_i}	Average number of relayed packets by CHs in the ring i
e_r	Energy density at the receiver antenna
e_s	Energy consumed during SS process

e_0	Energy consumed by the node in the running mode
e_i	Average of energy consumed by a node in ring i
L	Lifespan of the network
e_{int}	Initial stored energy of the node
e_{max}	Maximum energy consumed by a node within a cycle
D	Density of the nodes in the network
K_{opt}	Optimal number of rings
e_{opt}	Optimal consumed energy
P_c	Collision probability with PU during overall data gathering
P_{c_i}	Probability of collision with PU in the ring i
τ_s	Time slot of the PU activity Markov model
T_{a_i}	Average time of transmission activity of CH in ring i
T_{CN}	Period of data packet transmission from one single CN to its CH
τ_{T_i}	Average time of collecting data of a CH in ring i from one of its previous CHs
V_i	Average number of previous CHs in the ring $i + 1$ of a CH in the ring i
T_P	Period of a single CH packet transmission
m_{p_j}	Number of physical zones in a physical ring j

Appendix A. Proof of Equation (20)

We assume that our CRSN contains P equal physical zones, as shown in Figure A1. Each zone contains an equal number of CHs collecting similar physical data packets. We can divide our network into k_p physical rings, where the smallest ring is divided into three physical zones. Hence, the area of one physical zone is:

$$A_p = \frac{\pi}{3} \left(\frac{R}{k_p} \right)^2 \quad (A1)$$

The number of the physical zone P in the overall network is:

$$P = \frac{\pi R^2}{\frac{\pi}{3} \left(\frac{R}{k_p} \right)^2} = 3k_p^2 \quad (A2)$$

where R is the radius of overall CRSN. Similarly to Equation (1), the area of the i th physical ring is:

$$P_{A_i} = \pi \left(\frac{R}{k_p} \right)^2 (2i - 1) \quad (A3)$$

where $i \in [1, k_p]$. The number m_{p_j} of physical zones in physical ring j is equal to:

$$m_{p_j} = \frac{P_{A_i}}{A_p} = 3(2i - 1) \quad (A4)$$

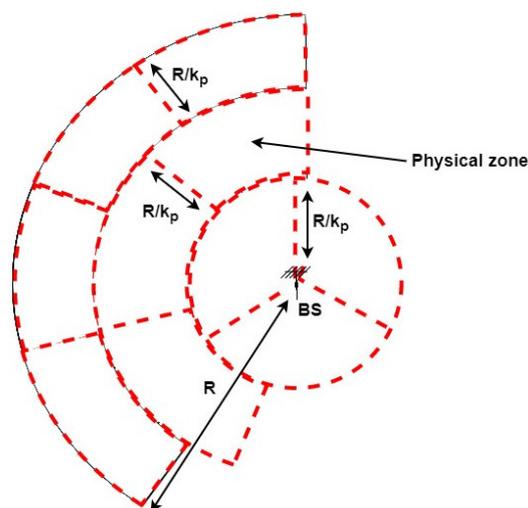


Figure A1. Physical zones in CRSN.

As shown in Figure A2, CHs in the SS ring i forward packets coming from physical zones located in the physical ring j . Obviously, in this example, we have CHs that forward one packet and others forward two packets coming from the same physical zone. In general, in the same SS ring, we have a set of CHs that forward n packets and others that forward $n + 1$ packets coming from the same physical ring. The average $av_{i,j}$ of the number of forwarding packets coming from a physical ring j by the CHs in SS ring i is calculated as:

$$av_{i,j} = \frac{1}{m_i} \sum_{t=1}^{m_i} \left(\left\lceil \frac{m_{p_j} t}{m_i} \right\rceil - \left\lfloor \frac{m_{p_j} (t-1)}{m_i} \right\rfloor \right) \tag{A5}$$

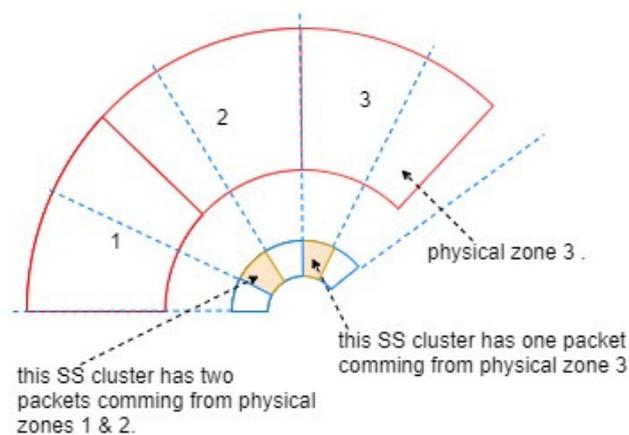


Figure A2. The forwarding packets coming from the physical ring j by the CHs in spectrum ring i .

The sum of average av_i of the number of forwarding packets coming from all physical zones by the CHs in spectrum ring i is:

$$av_i = \sum_{j=\lceil \frac{k_p}{k} i \rceil}^{k_p} \frac{1}{m_i} \sum_{t=1}^{m_i} \left(\left\lceil \frac{m_{p_j} t}{m_i} \right\rceil - \left\lfloor \frac{m_{p_j} (t-1)}{m_i} \right\rfloor \right) \tag{A6}$$

where $j = \lceil \frac{k_p}{k} i \rceil$ is the index of the physical ring containing the current SS ring i .

Finally, on average, a CH located in the ring i in our CRSN spends τ_{T_i} to forwarding packets toward its next CH located in the ring $i + 1$ in the way to the BS:

$$\tau_{T_i} = \frac{T_P}{m_i} \sum_{j=\lceil \frac{k_p}{k} i \rceil}^{k_p} \sum_{t=1}^{m_i} \left(\left\lceil \frac{m_{p_j} t}{m_i} \right\rceil - \left\lfloor \frac{m_{p_j} (t-1)}{m_i} \right\rfloor \right) \quad (\text{A7})$$

Appendix B. Proof of Equation (21)

Let $a = m_{p_j}$ and $b = m_i$, then:

$$\sum_{i=1}^b \left(\left\lceil \frac{a}{b} i \right\rceil - \left\lfloor \frac{a}{b} (i-1) \right\rfloor \right) = a + b - (a \wedge b) \quad (\text{A8})$$

Therefore, we have:

$$(a \wedge b) = a + b - \sum_{i=1}^b \left(\left\lceil \frac{a}{b} i \right\rceil - \left\lfloor \frac{a}{b} (i-1) \right\rfloor \right) \quad (\text{A9})$$

Let $(a \wedge b) = c$, then $a = cM$ and $b = cN$ with $M \wedge N = 1$.

1. If $a = b$, then $(a \wedge a) = a$ and:

$$a + b - \sum_{i=1}^b \left(\left\lceil \frac{a}{b} i \right\rceil - \left\lfloor \frac{a}{b} (i-1) \right\rfloor \right) \quad (\text{A10})$$

$$= a + b - \sum_{i=1}^b ([i] - [i-1]) \quad (\text{A11})$$

$$= a + b - \sum_{i=1}^b 1 = a \quad (\text{A12})$$

2. If $a > b$, then we get:

$$a + b - \sum_{i=1}^b \left(\left\lceil \frac{M}{N} i \right\rceil - \left\lfloor \frac{M}{N} (i-1) \right\rfloor \right) = b - \sum_{i=1}^{b-1} \left(\left\lceil \frac{M}{N} i \right\rceil - \left\lfloor \frac{M}{N} i \right\rfloor \right) \quad (\text{A13})$$

As we can see, if $\frac{M}{N} i \in \mathbb{N}$, then $\left\lceil \frac{M}{N} i \right\rceil - \left\lfloor \frac{M}{N} i \right\rfloor = 0$

In this case, we have $i \in S = \{N, 2N, \dots, (c-1)N\}$; then, we get:

$$\begin{aligned} & b - \sum_{i=1}^{b-1} \left(\left\lceil \frac{M}{N} i \right\rceil - \left\lfloor \frac{M}{N} i \right\rfloor \right) \\ &= b - \left(\sum_{i=1}^{b-1} 1 - \text{card}(S) \right) \\ &= b - b + c = c \end{aligned} \quad (\text{A14})$$

3. If $b > a$, it can be proved similarly to the case of $a > b$.

Appendix C. Proof of Equation (26)

According to Equation (A2), if $k_p = 1$, that means that we have just three distinct physical zones, in this case:

$$N_{p_i} = \frac{1}{m_i} \sum_{j=1}^1 \left(m_{p_1} + m_i - (m_{p_1} \wedge m_i) \right) \quad (\text{A15})$$

where, according to Equation (A4), $m_{p_1} = 3$. We get:

$$N_{p_i} = \frac{1}{m_i} \sum_{j=1}^1 m_i = 1 \quad (\text{A16})$$

which means all CHs in the overall network relay just one packet. So, let us consider the case where $k_p \geq 2$. Hence, to demonstrate that $N_{p_1} > N_{p_i}, \forall i > 1$, we should prove that:

$$k_p^2 - \frac{1}{m_i} \sum_{j=\alpha}^{k_p} \left(m_{p_j} + m_i - (m_{p_j} \wedge m_i) \right) > 0 \quad (\text{A17})$$

where $\alpha = \lceil \frac{k_p}{k} i \rceil$. Having:

$$\frac{1}{m_i} \sum_{j=\alpha}^{k_p} \left(m_{p_j} + m_i \right) > \frac{1}{m_i} \sum_{j=\alpha}^{k_p} \left(m_{p_j} + m_i - (m_{p_j} \wedge m_i) \right) \quad (\text{A18})$$

Inequality (A17) can be bounded by inequality (A18). We should prove:

$$k_p^2 - \frac{1}{m_i} \sum_{j=\alpha}^{k_p} \left(m_{p_j} + m_i \right) > 0 \quad (\text{A19})$$

Hence:

$$\begin{aligned} & k_p^2 - \frac{1}{m_i} \sum_{j=\alpha}^{k_p} \left(m_{p_j} + m_i \right) \\ &= k_p^2 - \frac{1}{3(2i-1)} \sum_{j=\alpha}^{k_p} \left(3(2j-1) + 3(2i-1) \right) \\ &= k_p^2 - \frac{1}{(2i-1)} \sum_{j=\alpha}^{k_p} \left((2j-1) + (2i-1) \right) \\ &= k_p^2 - \frac{1}{(2i-1)} \left(k_p^2 + k_p(2i-1) + (1-\alpha)(2i+\alpha-2) \right) \\ &= k_p^2(2i-2) - k_p(2i-1) + (\alpha-1)(2i+\alpha-2) \end{aligned} \quad (\text{A20})$$

In the last equation, $(\alpha-1)(2i+\alpha-2) \geq 0$ since $1 \leq \alpha \leq k_p$ and $i > 1$. Finally, we have to prove that $k_p^2(2i-2) - k_p(2i-1) > 0$. For this, we solve the following equation:

$$k_p^2(2i-2) - k_p(2i-1) = 0 \quad (\text{A21})$$

Solving the equation above, we get:

$$\begin{aligned} k_p &= 0, \text{ unacceptable since } k_p \geq 2 \\ k_p &= \frac{2i-1}{2i-2} < \frac{3}{2}, \forall i > 0, \text{ unacceptable since } k_p \geq 2 \end{aligned}$$

Therefore, the last equation has the same sign as k_p^2 . Hence, Equation (A20) is always positive. Finally, the inequality (A17) is proved.

References

1. Nasser, A.; Al Haj Hassan, H.; Abou Chaaya, J.; Mansour, A.; Yao, K.C. Spectrum Sensing for Cognitive Radio: Recent Advances and Future Challenge. *Sensors* **2021**, *21*, 2408. [[CrossRef](#)] [[PubMed](#)]
2. Ahmad, W.S.H.M.W.; Radzi, N.A.M.; Samidi, F.; Ismail, A.; Abdullah, F.; Jamaludin, M.Z.; Zakaria, M. 5G technology: Towards dynamic spectrum sharing using cognitive radio networks. *IEEE Access* **2020**, *8*, 14460–14488. [[CrossRef](#)]
3. Matinmikko-Blue, M.; Yrjölä, S.; Ahokangas, P. Spectrum management in the 6G era: The role of regulation and spectrum sharing. In Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 17–20 March 2020; pp. 1–5.
4. Hong-jiang, L.; Qi, Z. Tradeoff between utilization and collision in cognitive radio. In Proceedings of the 2009 International Conference on Wireless Communications & Signal Processing, Nanjing, China, 13–15 November 2009; pp. 1–5.
5. Tripathi, S.B.; Shah, M.B. A dynamic opportunistic spectrum access MAC protocol for Cognitive Radio networks. In Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, 24–27 September 2014; pp. 1195–1198. [[CrossRef](#)]
6. Eljack, S.; Igbal, M.; Furong, W. A Multi Channel Cognitive MAC Protocol with Efficient Channel Reservation and Collision Avoidance Method. In Proceedings of the 2009 International Conference on Multimedia Information Networking and Security, Washington, DC, USA, 18–20 November 2009; pp. 115–119. [[CrossRef](#)]
7. Stevenson, C.R.; Chouinard, G.; Lei, Z.; Hu, W.; Shellhammer, S.J.; Caldwell, W. IEEE 802.22: The first cognitive radio wireless regional area network standard. *IEEE Commun. Mag.* **2009**, *47*, 130–138. [[CrossRef](#)]
8. Mroue, M.; Nasser, A.; Parrein, B.; Mansour, A.; Zaki, C.; Cruz, E.M. ESco: Eligibility score-based strategy for sensors selection in CR-IoT: Application to LoRaWAN. *Internet Things* **2021**, *13*, 100362. [[CrossRef](#)]
9. Seshukumar, K.; Saravanan, R.; Suraj, M. Spectrum sensing review in cognitive radio. In Proceedings of the 2013 International Conference on Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT), Tiruvannamalai, India, 7–9 January 2013; pp. 1–4.
10. Joshi, G.P.; Kim, S.W. A survey on node clustering in cognitive radio wireless sensor networks. *Sensors* **2016**, *16*, 1465. [[CrossRef](#)] [[PubMed](#)]
11. Rani, S.; Ahmed, S.H. *Multi-Hop Routing in Wireless Sensor Networks: An Overview, Taxonomy, and Research Challenges*; Springer: Singapore, 2015.
12. Al-Karaki, J.N.; Kamal, A.E. Routing techniques in wireless sensor networks: a survey. *IEEE Wirel. Commun.* **2004**, *11*, 6–28. [[CrossRef](#)]
13. Rizvi, S. Cooperative multipath routing in wireless sensor networks. In Proceedings of the 2016 19th International Multi-Topic Conference (INMIC), Islamabad, Pakistan, 5–6 December 2016; pp. 1–6.
14. Uddin, M.A.; Mansour, A.; Jeune, D.L.; Ayaz, M.; Aggoune, E.H.M. UAV-assisted dynamic clustering of wireless sensor networks for crop health monitoring. *Sensors* **2018**, *18*, 555. [[CrossRef](#)]
15. Egidius, P.M.; Abu-Mahfouz, A.M.; Ndiaye, M.; Hancke, G.P. Data aggregation in software-defined wireless sensor networks: A review. In Proceedings of the 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, Australia, 13–15 February 2019; pp. 1749–1754.
16. Harb, H.; Abou Jaoude, C.; Makhoul, A. An energy-efficient data prediction and processing approach for the internet of things and sensing based applications. *Peer-to-Peer Netw. Appl.* **2020**, *13*, 780–795. [[CrossRef](#)]
17. Zhao, H.; Zhou, W.; Gao, Y. Energy efficient and cluster based routing protocol for WSN. In Proceedings of the 2012 Eighth International Conference on Computational Intelligence and Security, Kuala Lumpur, Malaysia, 3–5 July 2012; pp. 107–111.
18. Li, L.; Jiang, X.; Zhong, S.; Hu, L. Energy balancing clustering algorithm for wireless sensor network. In Proceedings of the 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuhan, China, 25–26 April 2009; pp. 61–64.
19. Wei, X.; Lu, J.; Zhuang, Y.; Ling, X. Adaptive clustering routing optimization for wireless sensor networks. In Proceedings of the 2013 Ninth International Conference on Natural Computation (ICNC), Shenyang, China, 23–25 July 2013; pp. 1011–1015.
20. Raghuvanshi, A.; Tiwari, S.; Tripathi, R.; Kishor, N. GK clustering approach to determine optimal number of clusters for Wireless Sensor Networks. In Proceedings of the 2009 Fifth International Conference on Wireless Communication and Sensor Networks (WCSN), Allahabad, India, 15–19 December 2009; pp. 1–6.
21. Alvarado, G.; Bosquez, C.; Palacios, F.; Córdoba, L. Low-energy Adaptive Clustering Hierarchy protocol and optimal number of cluster head algorithm in a randomized wireless sensor network deployment. In Proceedings of the 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, India, 15–16 December 2017; pp. 1–4.
22. Rauniyar, A.; Shin, S.Y. A novel energy-efficient clustering based cooperative spectrum sensing for cognitive radio sensor networks. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 198456. [[CrossRef](#)]
23. Zhang, H.; Zhang, Z.; Dai, H.; Yin, R.; Chen, X. Distributed spectrum-aware clustering in cognitive radio sensor networks. In Proceedings of the 2011 IEEE global telecommunications conference-GLOBECOM 2011, Houston, TX, USA, 5–9 December 2011; pp. 1–6.
24. Yadav, R.N.; Misra, R.; Saini, D. Energy aware cluster based routing protocol over distributed cognitive radio sensor network. *Comput. Commun.* **2018**, *129*, 54–66. [[CrossRef](#)]

25. Majid, A.; Azam, I.; Waheed, A.; Zain-ul Abidin, M.; Hafeez, T.; Khan, Z.A.; Qasim, U.; Javaid, N. An energy efficient and balanced energy consumption cluster based routing protocol for underwater wireless sensor networks. In Proceedings of the 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, Switzerland, 23–25 March 2016; pp. 324–333.
26. Guan, K.; He, L.M. A novel energy-efficient multi-path routing protocol for wireless sensor networks. In Proceedings of the 2010 International Conference on Communications and Mobile Computing, Shenzhen, China, 12–14 April 2010.
27. Gulbahar, B.; Akan, O.B. Information theoretical optimization gains in energy adaptive data gathering and relaying in cognitive radio sensor networks. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 1788–1796. [[CrossRef](#)]
28. Yan, M.; Ji, S.; Cai, Z. Time efficient data aggregation scheduling in cognitive radio networks. In Proceedings of the 2013 IEEE Global Communications Conference (GLOBECOM), Montreal, QC, Canada, 25–28 September 2006; pp. 1185–1190.
29. Yan, M.; Han, M.; Ai, C.; Cai, Z.; Li, Y. Data aggregation scheduling in probabilistic wireless networks with cognitive radio capability. In Proceedings of the 2016 IEEE global communications conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016; pp. 1–6.
30. Chen, Q.; Cai, Z.; Cheng, L.; Gao, H. Low-latency data aggregation scheduling for cognitive radio networks with non-predetermined structure. *IEEE Trans. Mob. Comput.* **2020**, *20*, 2412–2426. [[CrossRef](#)]
31. Rostami, A.S.; Badkoobe, M.; Mohanna, F.; Hosseinabadi, A.A.R.; Sangaiah, A.K. Survey on clustering in heterogeneous and homogeneous wireless sensor networks. *J. Supercomput.* **2018**, *74*, 277–323. [[CrossRef](#)]
32. Liu, X. A survey on clustering routing protocols in wireless sensor networks. *Sensors* **2012**, *12*, 11113–11153. [[CrossRef](#)]
33. Wei, D.; Jin, Y.; Vural, S.; Moessner, K.; Tafazolli, R. An energy-efficient clustering solution for wireless sensor networks. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 3973–3983. [[CrossRef](#)]
34. Lempinen, J.; Manninen, M. *Radio Interface System Planning for GSM/GPRS/UMTS*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
35. Kang, K.; Park, J.; Cho, S.; Jeong, B.; Kim, Y.; Lim, H.; Im, G. Deployment and coverage of cognitive radio networks in TV white space. *IEEE Commun. Mag.* **2012**, *50*, 88–94. [[CrossRef](#)]
36. Akkaya, K.; Ari, I. In-network data aggregation in wireless sensor networks. In *Handbook of Computer Networks: LANs, MANs, WANs, the Internet, and Global, Cellular, and Wireless Networks*; Wiley Online Library: Hoboken, NJ, USA, 2007; Volume 2, pp. 1131–1146.
37. Mantri, D.; Prasad, N.R.; Prasad, R. Scheduled Collision Avoidance in wireless sensor network using Zigbee. In Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, 24–27 September 2014; pp. 2129–2134.
38. Rida, M.; Makhoul, A.; Harb, H.; Laiymani, D.; Barhamgi, M. EK-means: A new clustering approach for datasets classification in sensor networks. *Ad Hoc Netw.* **2019**, *84*, 158–169. [[CrossRef](#)]
39. Harb, H.; Makhoul, A.; Laiymani, D.; Jaber, A. A distance-based data aggregation technique for periodic sensor networks. *ACM Trans. Sens. Netw.* **2017**, *13*, 1–40. [[CrossRef](#)]
40. Malathi, L.; Gnanamurthy, R.; Chandrasekaran, K. Energy efficient data collection through hybrid unequal clustering for wireless sensor networks. *Comput. Electr. Eng.* **2015**, *48*, 358–370. [[CrossRef](#)]
41. Ghosh, C.; Cordeiro, C.; Agrawal, D.P.; Rao, M.B. Markov chain existence and hidden Markov models in spectrum sensing. In Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications, Galveston, TX, USA, 9–13 March 2009; pp. 1–6.
42. Cheng, R.H.; Huang, C. The impact of the transmission power range on energy consumption for wireless sensor networks. In Proceedings of the 2013 Fifth International Conference on Ubiquitous and Future Networks (ICUFN), Da Nang, Vietnam, 2–5 July 2013; pp. 77–81.
43. Nasser, A.; Mansour, A.; Yao, K.C.; Charara, H.; Chaitou, M. Spectrum sensing for full-duplex cognitive radio systems. In Proceedings of the International Conference on Cognitive Radio Oriented Wireless Networks, Poznan, Poland, 11–12 June 2019; pp. 363–374.