Received: 11 June 2021 | Revised: 31 October 2021 | Accepted: 6 January 2022

DOI: 10.1049/cmu2.12343

IET Communications

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

An energy-efficiency-adaptive clustering formation mechanism for the wireless sensor networks

Deyu Lin[1,2] | **Linghe Kong**[1] | **Chengkun Zhao**[2] | **Jiayi Gao**[2] | **Hao Ouyang**[2] | **Ziyuan Yang**[3] | **Zhiqiang Zhang**[4]

[1] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

[2] School of Software, Nanchang Univeristy, Nanchang, China

[3] College of Computer Science, Sichuan University, Chengdu, China

[4] School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK

Correspondence
Linghe Kong, Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.
Email: linghe.kong@sjtu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61962019; Natural Science Foundation of Shaanxi Province, Grant/Award Number: 2020NY-175

Abstract
Energy inequality caused by the process of cluster head election has a large influence on energy efficiency and the network lifetime of wireless sensor networks (WSNs). To this end, a novel concept of EI_{ec} is proposed to evaluate the equality degree of energy consumption. Related theorems for establishing the candidate set of cluster heads are proposed, with the aim of promoting energy equality in each cluster. Subsequently, a novel energy-efficiency-adaptive cluster formation mechanism based on economic (ECFE) theory is proposed and detailed. Finally, extensive experiments are carried out to assess its energy efficiency and the network performance by comparisons with the existing classic and latest intelligent clustering algorithms. The results indicate that ECFE improves not only the energy efficiency but also the network performance effectively.

1 | INTRODUCTION

HOW to extend the network lifetime of the wireless sensor networks (WSNs) poses a huge challenge for the academic and industry all over the world [1, 2]. To this end, extensive attention has been paid to the improvement of the energy efficiency recently, with the aim of prolonging the network lifetime [3]. Actually, the energy efficiency is positively correlated with the network lifetime of WSNs [3]. In general, the energy efficiency is affected by the "Hot Spot Problem" and spatial-temporal correlation [4]. Specifically, the "Hot Spot Problem" results in energy inequality for the whole network topology, while spatial-temporal correlation leads to unnecessary energy overhead [5, 6].

Consequently, most of the existing energy-efficient techniques focus on two emphases, that is, the reduction and the equality in energy consumption to alleviate the problems of energy inequality and unnecessary energy overhead resulted from the "Hot Spot Problem" and spatial-temporal correlation [7, 8]. Generally, they can be classified into energy-efficient medium access control, energy-efficient mobile node assistance scheme, clustering mechanism, and energy-efficient routing scheme respectively [1].

Clustering mechanism is widely utilized in the applications of WSNs [9]. It aims to improve the energy efficiency by means of reducing and balancing the energy consumption concurrently. To be specific, it groups all of the sensor nodes into separated clusters logically. As a result, the nodes in each cluster are classified into two categories, that is, cluster member (CM) and cluster head (CH) respectively. CH is elected according to some predefined metrics, such as the amount of residual energy, the distance from the node to the sink, and so on. CM propagates the original data to its CH, which performs simple aggregations on the original data to reduce the energy waste resulted from spatial-temporal correlation subsequently. In addition, the energy overhead for medium access controlling can be also alleviated via TDMA mode in clustering mechanism [6]. To be specific, all the CMs within the same cluster switch between working and sleeping modes to reduce the duty-cycle to cut down energy overhead further.

In clustering mechanism, the process of cluster formation is of great importance since it affects the energy distribution of the entire cluster [7]. Consequently, the distribution of CHs has a large influence on that of energy for the whole network. Since each CH needs to receive and forward all the data from both of its own cluster and outer clusters to the sink, it bears a much heavier energy burden than CMs. Therefore, it is reasonable to rotate the role of CH round by round [7]. Consequently, both of the selection and the rotation of CHs are important for the clustering mechanism. Recently extensive focuses have been put on clustering mechanisms [10]. In some researches, CH is elected randomly to balance the energy consumption among different nodes within the same cluster. In LEACH, the cluster head is selected based on a random number [11]. Similarly, the nodes take turn to be CH round-by-round for a chain-shaped network topology in PEGASIS [12]. In some other scenarios, the metrics for CH determination are predefined to control the network delay to a certain extent. For example, a double-threshold mechanism was presented in TEEN to meet the demand of some hard real-time applications [13]. In addition, some interdisciplinary algorithms, such as the game theory [14, 15], the fuzzy logic theory [16, 17, 18], the simulated annealing algorithm [19], the particle swarm optimisation algorithm [20], or the multi-objective optimisation algorithm [21] etc., were proposed to improve the energy efficiency in WSNs. In fact, most of the problems in WSNs can be formulated into the multi-objective problem, with the aim of taking several factors concurrently [22, 23]. Finally, some other clustering mechanisms which belongs to the cross-layer algorithm, such as those combined with the transport layer [24–26], the MAC layer [27, 28] etc., were also proposed.

The clustering mechanisms listed above are able to improve the energy efficiency and prolong the network lifetime of WSNs to some extent. However, none of them takes the relationship between each individual and the network into consideration. In fact, it is a crucial factor for the energy equality during the process of cluster formation. However, to the best of our knowledge, all the existing clustering mechanisms ignore it, let alone the systematical analysis. Different from them, the relation between each sensor node and the corresponding cluster is taken into consideration through the systematical methodology in this paper, with the aim of alleviating energy inequality within each cluster. According to Appendix A, the theory of social welfare is adopted to control the process of CH selection in this paper, with the aim of promoting energy equilibrium in the whole network topology.

A novel energy-efficiency-adaptive cluster formation mechanism based on Economic (ECFE) theory to improve the energy efficiency within each cluster in this paper. In ECFE, the cluster head is elected according to both of the residual energy of sensor nodes and the equality degree of energy consumption simultaneously.

Specifically, the contributions of this paper are listed as follows.

1. A novel concept of EI_{ec} is proposed to realize energy equilibrium within each cluster. To be specific, it presents a novel concept of arithmetical means of squared Euclidean distance (SED) to assess energy equilibrium within each cluster.
2. Theorems for determining the optimal CH and the value of SED are proposed and proven in detail. As a result, the selection zone for the candidate can be established precisely. Meanwhile, the value of SED is determined to select the optimal CH subsequently.
3. Extensive simulations are conducted to evaluate ECFE's performance with regard to the energy efficiency and other network performance. Comparisons with the classic and latest intelligent clustering techniques under the equivalent evaluation metrics have verified its effectiveness.

The rest of this paper is organized as follows. Section 2 presents the preliminaries in details. Section 3 proposes the relevant concepts and theorems. Subsequently the novel mechanism of ECFE is proposed in detail in Section 4, and it is evaluated through extensive simulations in Section 5. Finally we draw the conclusions and point out some potential research directions in the future.

2 | PRELIMINARIES

In this section, the first-order radio model and the network topology are presented firstly. Subsequently, the related assumptions and notations utilized in this paper are described.

2.1 | First-order radio model and network topology

2.1.1 | First-order radio model

In this paper, we adopt the first-order radio model to quantify the energy expenditure for data communication [9]. To be specific, the energy consumption for data transmission and reception are listed respectively as expressions (1), (2).

$$e_{tx} = k(E_{elec} + \varepsilon_{amp} \cdot d^{\alpha}) \tag{1}$$

$$e_{rx} = kE_{elec} \tag{2}$$

where d denotes the transmission range, k the size of the packet, E_{elec} the energy consumption in the transmitter or receiver circuit, ε_{amp} the transmitter amplifier respectively. Finally, α is the propagation loss exponent. Its value depends on the transmission model. To be specific, there are two kinds of transmission model, namely the free space model and the multipath fading model. α is 2 for the free space model, and increases to be 4 for the latter [7].

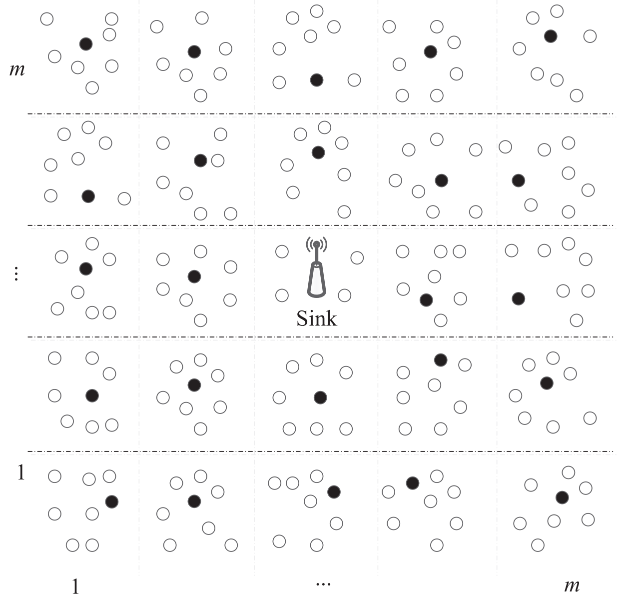

FIGURE 1 Network topology utilized in this paper

2.1.2 | Network topology

Assume a network topology with an arbitrary shape, which is divided into m^2 grids. The Sink locates at the centre of the network topology. The length of side for each gird is required to meet the following condition so as to control the transmission overhead [7]

$$\sqrt{5}a \leq d_{thre} \&\& \ d_{thre} = 87.7 \tag{3}$$

where d_{thre} denotes the threshold of transmission range for the free space model. As a result, the parameter of α is made to equal 2 with the establishment of condition (3).

Each grid is regarded as a single cluster, which contains CMs and CH as shown in Figure 1. Besides, the grid containing the sink does not select out any CHs to alleviate the "Hot Spot Problem."

2.2 | Related assumptions and notations

In this section, we present some related assumptions and notations for convenience. All of the notations are listed in Table 1 for the sake of brevity.

Each node is powered by the battery with the same amount of initial energy. On the contrary, the sink is supposed to be unlimited in energy supply and processing capacity.

Each sensor node keeps static and is aware of the location of both itself and its neighbours [29, 30]. The position information can be adopted to elect the optimal CH. Besides, the distribution of nodes follows the uniform distribution [25, 31]. Besides, the processes of inter-cluster routing construction and data acquisition are beyond the scope of this paper. Actually, they follow the same way of our previous work [13].

TABLE 1 Summary of related symbols and corresponding meanings

Symbols	Meanings
S	State of the sensor node
g_c	Position of the central grid
T	The cycle of rotation
ε	Energy inequality aversion parameter
M	Length of the side of the network topology
μ	Data generation rate of sensor node
a	The length of each grid

Each node can control the communication range adaptively to cut down the energy overhead as much as possible. In fact, it can obtain the distance from itself to its neighbours according to the received signal strength indicator (RSSI) concisely [32].

3 | CONCEPT AND THEOREMS CONCERNING THE PROCESS OF CLUSTER HEAD SELECTION

In this section, a novel concept of EI_{ec} is presented to evaluate the influence of the potential CH on energy efficiency firstly. Subsequently, the related theorems for the candidate establishment are proposed and proven at length. Finally, the details about the process of CH determination are illustrated.

3.1 | Energy equality index for data transmission

Definition 4.1. Equality Index for energy consumption (EI_{ec}) is defined here to assess the degree of energy equality for the whole network topology. To effectively reflect the balance condition, it follows the same form of Atkinson's inequality measure [37].

According to our previous work [33], Atkinson's inequality measure is adopted to penalize inequality. As a result, the concept of EI_{ec} can be mathematically expressed by the following equation

$$EI_{ec} = \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{d_{i-CH}^2}{\overline{d^2}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \tag{4}$$

where $\overline{d^2}$ denotes the arithmetical means of squared Euclidean distance (SED) from all the sensor nodes to the CH within the same cluster, and ε the inequality aversion parameter [37] which is adopted to penalize the energy inequality resulted from data transmission. As for the process of CH election, the node

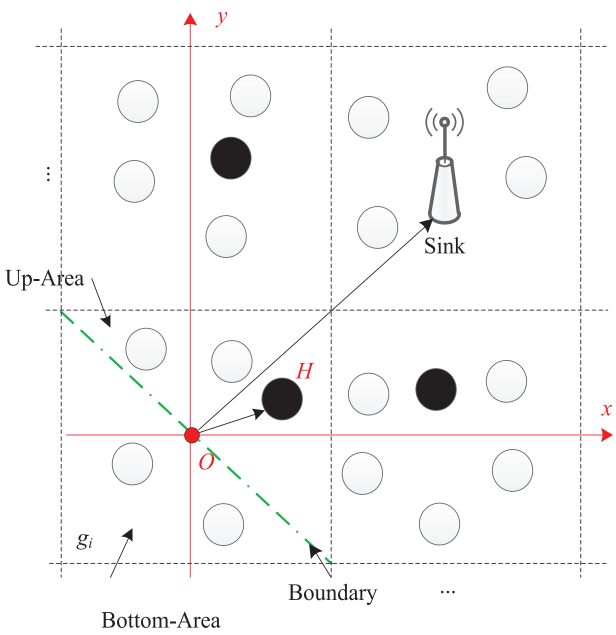

FIGURE 2 Illustration of electing the optimal candidate for CH in the next round

is more likely to be penalized with the degradation of energy equality. Since most of the energy is depleted for data transmission, it can be easily concluded that EI_{ec} is beneficial to bringing in energy equilibrium if it is close to 1.

3.2 | Theorems for candidate set construction

In this section, the theorem for candidate set construction CHs in the next round is presented in detail. The residual energy and the distance from the candidate to the sink are taken into consideration with the aim of controlling the energy overhead and energy equality simultaneously.

Theorem 4.1. *The optimal candidate for the CH in the next round needs to meet the following conditions,*

$$\begin{cases} Max.\ E_{re}^{i}\,\overrightarrow{OS}\cdot\overrightarrow{OH_i}\,, & if\ \ \overrightarrow{OS}\cdot\overrightarrow{OH_i} > 0 \\ Max.\ E_{re}^{i}\Big/|\overrightarrow{OH_i}|\,, & if\ \ \overrightarrow{OS}\cdot\overrightarrow{OH_i} = 0 \\ Max.\ \overrightarrow{OS}\cdot\overrightarrow{OH_i}\Big/E_{re}^{i}, & if\ \ \overrightarrow{OS}\cdot\overrightarrow{OH_i} < 0 \end{cases} \tag{5}$$

where $\overrightarrow{OS}$, $\overrightarrow{OH_i}$ denote the vector from the central point O *to the Sink, and that from* O *to the position of the possible CH respectively. E_{re}^{i} denotes the residual energy of node sn_i.*

Proof: As shown in Figure 2, suppose an arbitrary grid g_i lies in the network topology. Establish a coordinate system with the position O as the origin. Grid g_i is divided into three different parts by a green dash line, up-area, bottom-area, and boundary respectively. The green dash line, which is perpendicular to the line from O to the sink, is regarded as the boundary. According to the definition of energy efficiency [1], it is necessary to take the residual energy and the communication overhead into consideration concurrently when electing CH [7]. The inner product of vectors $\overrightarrow{OS}$ and $\overrightarrow{OH_i}$ can reflect the communication range clearly. In fact, the smaller the inner product, the closer the distance from the candidate to the sink. Therefore, the inner product is utilized to control the energy expenditure for the candidate because the communication overhead depends on the transmission range. For the node in the up-area, the following condition is met,

$$\overrightarrow{OS}\cdot\overrightarrow{OH_i} > 0 \tag{6}$$

To effectively control the residual energy and the communication consumption simultaneously, the following expression needs to be established,

$$Max.\ E_{re}^{i}\,\overrightarrow{OS}\cdot\overrightarrow{OH_i}\,,\quad if\ \ \overrightarrow{OS}\cdot\overrightarrow{OH_i} > 0 \tag{7}$$

For the node on the boundary, it is easily obtained that the inner product of vectors $\overrightarrow{OS}$ and $\overrightarrow{OH_i}$ is zero

$$\overrightarrow{OS}\cdot\overrightarrow{OH_i} = 0 \tag{8}$$

Therefore, the following expression is presented here to elect the candidate,

$$Max.\ E_{re}^{i}\Big/|\overrightarrow{OH_i}|,\quad if\ \ \overrightarrow{OS}\cdot\overrightarrow{OH_i} = 0 \tag{9}$$

where $|\overrightarrow{OH_i}|$ denotes the magnitude of vector $\overrightarrow{OH_i}$. Obviously, the node with maximum value of $E_{re}^{i}/|\overrightarrow{OH_i}|$ possesses the largest value of residual energy and locates closest to the possible CH concurrently.

Finally, as for the node lying in bottom-area, the inner product of vectors $\overrightarrow{OS}$ and $\overrightarrow{OH_i}$ is negative. The larger the inner product, the lower energy the candidate needs for communication. Besides, with the rise of its residual energy, the value of $\overrightarrow{OS}\cdot\overrightarrow{OH_i}/E_{re}^{i}$ increases. Therefore the following condition can be obtained to elect a proper CH,

$$Max.\ \overrightarrow{OS}\cdot\overrightarrow{OH_i}\Big/E_{re}^{i},\quad if\ \ \overrightarrow{OS}\cdot\overrightarrow{OH_i} < 0 \tag{10}$$

On the basis of Theorem 4.1, the candidate for the optimal CH in the next round can be determined. As a result, the candidate set C is established as follows

$$C = \{c_{up},\ c_{by},\ c_{bm}\} \tag{11}$$

where c_{up}, c_{by}, and c_{bm} denote the candidates chosen from up-area, boundary, and bottom-area respectively.

3.3 | CH determination based on expected energy efficiency welfare

In this section, the concept of Energy Efficiency Welfare, which is as shown in Appendix B, is adopted to determine the optimal CH. Besides, the theorem for determining the arithmetical means of SED, namely the value of $\overline{d^2}$, is proposed and proven at length. Finally, the process of expected energy efficiency welfare determination is detailed based on the value of the arithmetical means of SED.

Theorem 4.2. *For an arbitrary grid g in which the distribution density of sensor nodes is λ, the distance from the CH candidate to the central point of this grid is d_{CH-O}, then the arithmetical means of SED can be obtained as follows*

$$\overline{d^2} = \frac{\frac{1}{6}a^2 + d^2_{CH-O}}{\lambda} \tag{12}$$

Proof: As shown in Figure 3, assume the location of the possible CH is denoted as (L_x^{ch}, L_y^{ch}). Establish a coordinate system with the central point as the origin, then SED from other nodes to the CH candidate is obtained as follows,

$$\sum d_i^2 = \lambda \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{a}{2}}^{\frac{a}{2}} \left[\left(x - L_x^{ch}\right)^2 + \left(y - L_y^{ch}\right)^2 \right] dxdy \tag{13}$$

According to the assumption in Section 3, the total number of sensor nodes in grid g can be obtained as below,

$$N_g = \lambda a^2 \tag{14}$$

where N_g is the number of sensor nodes in grid g.

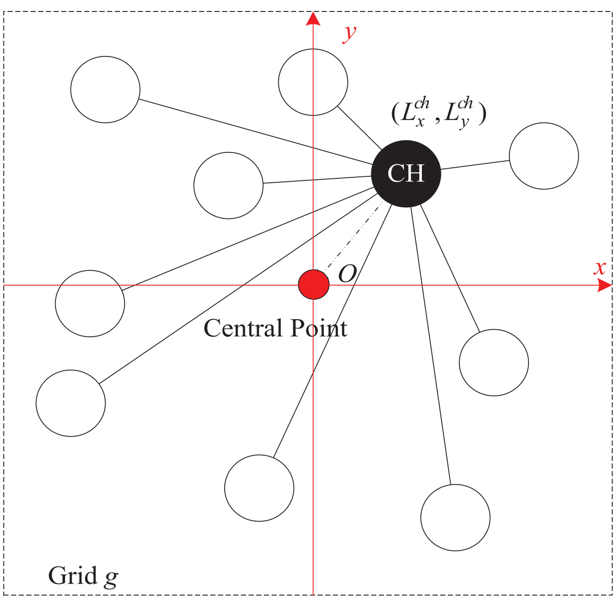

FIGURE 3 Determination of the arithmetical means of SED

Based on expressions (13), (14), the arithmetical means of SED is obtained as follows,

$$\overline{d^2} = \frac{\frac{1}{6}a^2 + \left(L_x^{ch}\right)^2 + \left(L_y^{ch}\right)^2}{\lambda} \tag{15}$$

Obviously, we can get the arithmetical means of SED as below finally,

$$\overline{d^2} = \frac{\frac{1}{6}a^2 + d^2_{CH-O}}{\lambda}$$

With the arithmetical means of SED, the expected energy efficiency welfare can be established. In a summary, the node in set C calculates its expected residual energy firstly. Subsequently, it starts to compute the expected energy efficiency welfare according to Definition 4.2 (as shown in Appendix B).

Suppose an arbitrary node c_j in set C, it estimates the expected residual energy of all the other nodes within gird g according to the following expression,

$$\hat{E}_{re}^k = E_{re}^k - \mu T \varepsilon_{amp} {d_{k-c_j}}^2, \quad k \in SN - \{c_j\}. \tag{16}$$

where $\hat{E}_{re}^k$ denotes the expected residual energy of sensor node sn_k and d_{k-c_j} is the distance from sn_k to candidate c_j.

Subsequently, c_j estimates the expected energy efficiency welfare $E^2\hat{W}(c_j)$ from its own perspective based on the following expression,

$$E^2\hat{W}(c_j) = E_{re}^k \cdot \left[\frac{1}{n} \sum_{k=1}^{n} \left(\frac{d^2_{k-c_j}}{\overline{d^2}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \tag{17}$$

where n denotes the number of the possible CMs of candidate c_j. Likewise, all the other candidates in set C obtain the expected energy efficiency welfare $E^2\hat{W}(c_l)$, $c_l \in C - \{c_j\}$ from their own standpoint.

Once all the candidates in set C obtain the expected energy efficiency welfare, they begin to compete for the role of CH in the next round. Finally, the node with the highest expected energy efficiency welfare is elected as CH.

4 | ENERGY-EFFICIENCY-AWARE CLUSTER FORMATION MECHANISM

In this section, the novel cluster formation mechanism of ECFE is proposed. After a brief introduction, the details of ECFE are presented [34].

4.1 | Introduction of ECFE

On the whole, ECFE consists of three main phases, namely, candidate set construction phase, expected energy efficiency welfare determination phase, and optimal CH election phase respectively. In ECFE, both of the residual energy and the equality degree of energy consumption are taken into consideration concurrently. Once the cluster is formed, the processes of routing determining and data acquisition start. According to the assumptions expounded in Section III, these processes are beyond the scope of this paper. Therefore, the emphasis is put on the process of cluster formation in the following paragraphs.

4.2 | Details of ECFE

In this section, the three phases of ECFE are illustrated in a detailed way. In fact, we will focus on the process of candidate set construction and the optimal CH determination.

4.2.1 | Candidate set construction

In the initial phase of ECFE, all of the sensor nodes need to determine which candidate zones they belong to according to their location based on Theorem 4.1. Each sensor node calculates the inner product of vectors $\overrightarrow{OS}$ and $\overrightarrow{OH_i}$ according to Theorem 4.1. On the basis of the value of the inner product, each sensor node is able to identify its belonging candidate zone.

Subsequently, the node in each zone select out its own candidate according to Theorem 4.1 independently. Take the nodes in up-area for example, the sensor which brings in maximum value of $E_{re}^{i}\overrightarrow{OS}\cdot\overrightarrow{OH_i}$ is elected as the CH candidate in the next round. Likewise, the candidates in the other zones can be determined simultaneously.

Finally, the selected nodes from all the zones constitute the candidate set C. In addition to the residual energy, the energy consumption of candidates for transmission has also been considered during the process of candidate selection according to Theorem 4.1. Consequently, the candidate is more beneficial to the improvement of energy efficiency than others in the same zone.

4.2.2 | Expected energy efficiency welfare determination

In this phase, the node in candidate set C obtains the arithmetical means of SED firstly, which is utilized to determine the value of EI_{ec}. To this end, each node in candidate set C calculates the value of $\overline{d^2}$ on the basis of expression (12). Subsequently, it is able to estimate the expected residual energy of other nodes within the same cluster except itself. At the same time, it estimates the expected energy efficiency welfare according to expression (17) with the results of the expected residual energy and the arithmetical means of SED. Likewise, the other candidates calculate the expected energy efficiency welfare in the similar way independently. Once all the candidates in set C obtain the value of $E^2\hat{W}(c_j)$, they are able to select out the optimal CH in the next round.

4.2.3 | Optimal cluster head election

In this phase, all the candidates in set C compete with each other to campaign for the role of CH in the next round. Assume the current candidate is c_j. Once candidate c_j obtains the value of the expected energy efficiency welfare, it modifies its own state S to be 1 immediately. Subsequently it broadcasts a message including the basic information, such as ID, S, etc., and the expected energy efficiency welfare to all the other candidates.

On receiving the broadcasts from other candidates, candidate c_j compares its own expected energy efficiency welfare with what is included in the broadcasts to determine who is the optimal CH. To be specific, if its own $E^2\hat{W}(c_j)$ is smaller, it accepts the transmitter as the optimal CH, then withdraws from election by means of setting the value of S to be 0. In addition, it needs to modify its record concerning the CH in its memory. Otherwise, it only discards the broadcast and keeps its state S to be 1.

When there are not any broadcasts flooding in the grid, the optimal CH is selected out successfully. All the other nodes failing to CH election need to join the cluster head. Subsequently, the cluster head informs CMs of its role and waits for their JOIN message. As for the cluster member, it simply sends out the JOIN message to the CH to form a cluster in the end [13].

Once all the relevant CHs are elected, the process of intercluster routing decision-making starts. Subsequently CMs begin to acquire data and all the data collected in each cluster are transmitted to the sink via a multi-hop transmission pattern finally.

5 | EXPERIMENTS AND RESULTS ANALYSIS

In this section, the experiment settings are described firstly. Subsequently, the relevant metrics utilized in this paper are defined. Finally, the results and analysis are detailed.

5.1 | Experiment settings

ECFE belongs to a kind of Clustering mechanisms which aim to promote energy equilibrium during the process of cluster formation. Different from the existing clustering mechanisms, the relation between each sensor node and the whole network is taken into account based on the energy efficiency welfare. Besides, it is for the first time that CH selection is controlled by the systematical social theory. To verify its superiority over traditional clustering mechanisms, the classic clustering algorithms, such as TEEN, and PEGASIS, are adopted as the baseline for evaluation. Besides, to verify its merit over interdisciplinary clustering mechanisms, two intelligent clustering algorithms, that is, LEACH-ERE [17] and EIRNG which aims to improve the energy efficiency in inter-cluster routing determination based on the game theory [35], are adopted for comparisons in this paper.

In this section, the performance of ECFE is evaluated through simulations. In the experiment, all the sensor nodes are

TABLE 2 Some important parameters and their values in simulation

Parameters	Values
a	30 m
m	{3, 5, 7, 9, 11}
g_c	{(2,2), (3,3), (4,4), (5,5), (6,6)}
ε	2
E_{elec}	50 nJ/bit
ε_{amp}	13 pJ/bit/m^2
T	50 s

scattered in an $ma \times ma$ square area. The distribution of sensor nodes follows the uniform distribution. The parameters of m and a are described in section 3. Let a equal 30, and m vary from 3 to 11 with the step of 2. The experiments are grouped into five sets and extensive simulations are conducted on each set to assess the energy efficiency comprehensively. In addition, The initial energy of each sensor node is fixed to be 2 J. For the sake of brevity, the values of some important parameters utilized in the simulation are listed in Table 2. In addition, the total number of rounds in the simulation is set to be 80.

5.2 | Metrics utilized in the experiments

In order to evaluate the energy efficiency of ECFE objectively, we firstly define some related metrics in the following.

The network lifetime is one of the main metrics for ECFE's effectiveness evaluation. In general, the definition of network lifetime depends on the related applications. In this paper, the following three indicators are adopted to evaluate the network lifetime.

Time until the first node dies (FND): It denotes the time quantum until the first node has exhausted its energy. For the application with a high demand on the reliability, such as endangered species tracking, military monitoring etc. Consequently, it is crucial to evaluate the metric of FND.

Time until half of the nodes die (HND): It denotes the time period until half of the nodes have used up their energy. Actually, it reflects the coverage rate of the network to some extent.

Time until the last node dies (LND): It represents the duration until all the sensor nodes have exhausted their energy.

The indicators of Network Lifetime above can be adopted to measure the energy efficiency indirectly. In addition, some metrics are defined to assess the energy efficiency of WSNs intuitively.

Throughput against energy consumed: It denotes the throughput of the sink when a certain amount of the total energy has been used up. It takes both the amount of data and the energy expenditure into consideration simultaneously, therefore it reflects the energy efficiency intuitively.

Average of residual energy: It reflects the average residual energy for all the nodes in the process of simulation. Since the energy overhead for transmission is taken into account in ECFE, it is adopted to evaluate the energy depletion rate.

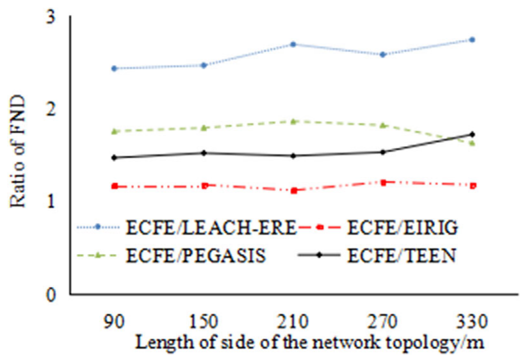

FIGURE 4 The ratios of our proposal to other algorithms in terms of FND

Approximate value of energy efficiency welfare: It reflects the approximate value of energy efficiency welfare through the considerations of both average residual energy and the variance of residual energy for all the nodes during the process of experiment. To be specific, it is defined mathematically as follows,

$$AV_{eew} = \overline{E_{re}} \cdot (1 - E_{re}^{va}) \tag{18}$$

where AV_{eew}, $\overline{E_{re}}$, and E_{re}^{va} denote the approximate value of energy efficiency welfare, the average of residual energy, and the variance of residual energy respectively. According to the definition, the energy efficiency of ECFE can be assessed objectively.

5.3 | Results analysis

In this section, the ratios of ECFE to other algorithms in terms of the Network Lifetime and the throughput are analysed with different values of parameter m. While the results analysis on other metrics, such as the throughput against the energy consumed, average residual energy, etc., is conducted with parameter m set to be 3.

The comparisons on the ratio of our proposal to other algorithms in terms of FND are presented in Figure 4. As shown in Figure 4, the ratios vary with the change of parameter m. All the ratios are larger than 1, which means ECFE can bring in a longer network lifetime of WSNs compared with other clustering algorithms with respect to FND. In addition, it can be also easily obtained from Figure 4 that the ratios kept steady in spite of the growth of parameter m. Consequently, the conclusion that ECFE has a good scalability in energy efficiency can be drawn.

Figures 5 and 6 show the comparisons on the ratios of ECFE to others in terms of HND and LND respectively. As shown in Figures 5 and 6, it is clear that the ratios are much larger than 1, which means the network lifetime is the largest for ECFE. Take the comparisons on HND as an example, the ratios of ECFE/LEACH-ERE, ECFE/EIRNG, ECFE/PEGASIS, and ECFE/TEEN are 2.43, 1.22, 1.42, and 1.55 when parameter m is set to be 3 respectively. Since the energy inequality resulted from the process of CH election is alleviated through the concept of EI_{ec}, the energy efficiency can be improved effectively.

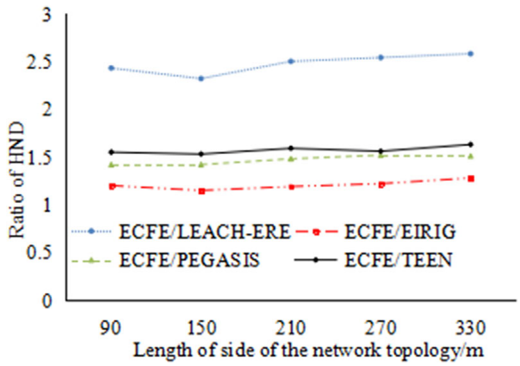

FIGURE 5 The ratios of our proposal to other algorithms in terms of HND

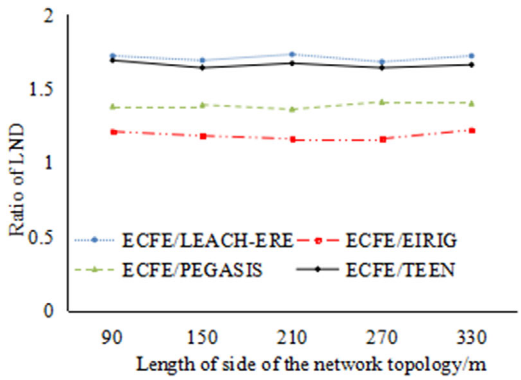

FIGURE 6 The ratios of our proposal to other algorithms in terms of LND

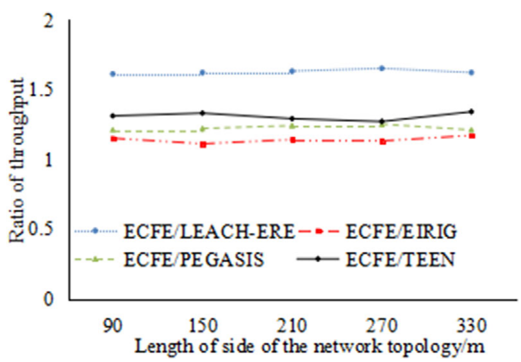

FIGURE 7 The ratio of our proposal to other algorithms in terms of the throughput

According to the relation between energy efficiency and network lifetime, the former is extended with the improvement of the latter accordingly.

Figure 7 shows the ratios of ECFE to others in terms of the throughput. Specifically, the throughput of ECFE increases by 63.12%, 14.20%, 24.32%, and 29.46% compared with LEACH-ERE, EIRNG, PEGASIS, and TEEN respectively when parameter m is set to be 7. In addition, it can also be obtained that ECFE has a largest effect on the throughput compared with LEACH-ERE with the ratio of 1.65. The energy inequality

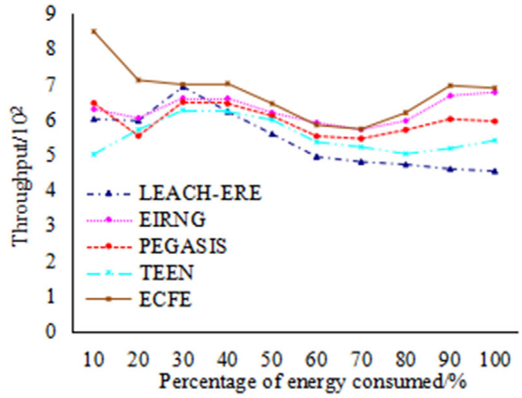

FIGURE 8 The comparisons on the throughput when a certain amount of energy has been exhausted and parameter m is set to be 3

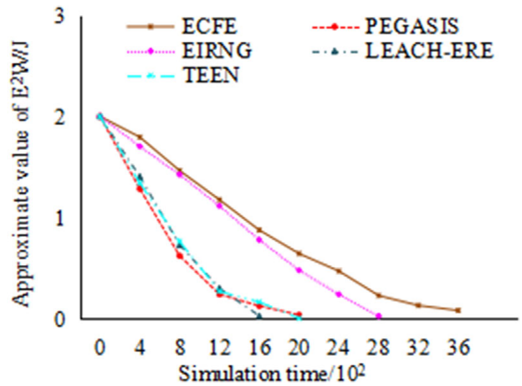

FIGURE 9 The approximate value of energy efficiency welfare of all the algorithms when parameter m is set to be 3

resulted from the process of CH election is effectively alleviated through the concept of EI_{ec}, the energy efficiency can be largely improved accordingly. The network lifetime can be extended effectively for ECFE, as a result, the throughput is relatively larger than that of others by the end of the simulation. As a summary, with the extension of network lifetime in ECFE, the total throughput of the sink rises up accordingly.

Figure 8 shows the comparisons on the throughput when a certain amount of energy has been exhausted and the parameter m is set to be 3. As shown in Figure 8, the throughput of ECFE under a specific energy consumption is much larger than those of others, which means more data can be collected by the sink when ECFE is applied under a fixed energy budget. In addition, note that when more than half of the total energy has been used up, the throughput of ECFE increases more markedly compared with others. Specifically, when 140 J of the total energy is used up, the throughput of the sink is 31.07%, 3.86%, 8.53%, and 23.23% larger than that of LEACH-ERE, EIRNG, PEGASIS, and TEEN respectively. Therefore the conclusion that the energy efficiency of ECFE is the highest compared with others can be drawn.

Figure 9 shows the approximate value of energy efficiency welfare of all the algorithms. As shown in Figure 9, the average residual energy of ECFE is 135.97%, 2.81%, 100.69%, and

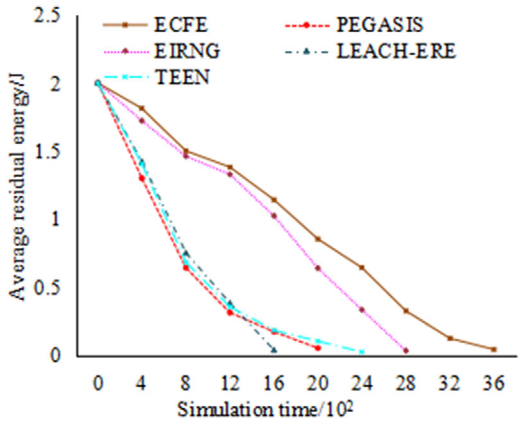

FIGURE 10 The average of residual energy of sensor nodes for all of the algorithms when parameter m is set to be 3

91.92% larger than that of PEGASIS, EIRNG, LEACH-ERE, and TEEN when the simulation time is 800 respectively. In addition, the curve of ECFE last for longest by comparisons with others, which also means it promotes a longer network lifetime. As an important component for ECFE, the concept of EI_{ec} is helpful to the improvement of energy equality among different nodes within each cluster.

Figure 10 shows the average residual energy of sensor nodes for all the algorithms when parameter m is set to be 3. As shown in Figure 10, the average of residual energy is the largest compared with others. At the time of 800, the ratios of ECFE to PEGASIS, EIRNG, LEACH-ERE, and TEEN are 134.17%, 2.73%, 99.70%, and 120.18% respectively. Besides, it keeps steady with the increase of the simulation time. The conclusion that ECFE can achieve a slower energy depletion rate can be drawn. Obviously it is beneficial to the improvement of energy efficiency and the extension of network lifetime for WSNs.

6 | CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The energy inequality resulted from cluster formation is analysed and solved in this paper. A novel concept of EI_{ec} is proposed to evaluate the equality degree of energy consumption. Besides, theorems for establishing the candidate set of CHs is proposed, with the aim of promoting energy equality within each cluster. Subsequently, a novel energy-efficiency-aware cluster formation mechanism of ECFE is proposed and detailed. Finally, extensive simulations are carried out to evaluate its performance. The results analysis indicates that ECFE can improve the energy efficiency and extend the network lifespan of WSNs effectively.

The energy efficiency of WSNs can be improved by means of controlling the process of cluster formation to some extent. However, most of the existing clustering mechanisms are based on the premise that the sensor node cannot be replenished once deployed. Recently, the emergence of the simultaneous wireless information and power transfer (SWIPT) technology makes it possible to recharge the sensor node with the ambient RF signals [36]. To be specific, the receiver can achieve information decoding and energy harvesting concurrently through SWIPT. However, combination with SWIPT technology also induces some new challenges for the traditional clustering mechanisms. For example, the allocation of transmission power and the value of power splitting ratio play a large influence on the process of CH election [37]. Therefore, our attention will be paid to the clustering mechanism for the WSNs combined with SWIPT in the future.

FUNDING INFORMATION

This paper was supported in part by the National Natural Science Foundation of China (61962019), and in part by Natural Science Foundation of Shaanxi Province (2020NY-175).

CONFLICT OF INTEREST

All authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Deyu Lin https://orcid.org/0000-0003-1400-4769

REFERENCES

1. Yetgin, H., Cheung, K.T.K., El-Hajjar, M., Hanzo, L.H.: A survey of network lifetime maximization techniques in wireless sensor networks. IEEE Commun. Surv. Tutorials 19(2), 828–854 (2017)
2. Lee, J., Ekici, E.: Sequential sensor selection and access decision for spectrum sharing. IEEE Trans. Aero. Electr. Syst. 55(3), 1062–1074 (2019)
3. Lin, D., Wang, Q., Min, W., Xu, J., Zhang, Z.: A survey of the energy efficient strategies in wireless sensor networks. ACM Trans. Sens. Netw. 17(1), 1–48 (2020)
4. Wang, F., Liu, J.: Networked wireless sensor data collection: Issues, challenges, and approaches. IEEE Commun. Surv. Tutorials 13(4), 673–687 (2011)
5. Lin, D., Min, W., Xu, J.: An energy-saving routing integrated economic theory with compressive sensing to extend the lifespan of WSNs. IEEE Internet Thing 7(8), 7636–7647 (2020)
6. Liu, Y., Zhang, Q., Ni, L.: A general framework for spectrum sensing using dedicated spectrum sensor networks. ACM Trans. Sens. Netw. 15(1), 1–23 (2019)
7. Andreev, S., Dobre, C., Misra, P.: The internet of things and sensor networks. IEEE Commun. Mag. 57(12), 64 (2019)
8. George, R., Mary, T.A.J.: Review on directional antenna for wireless sensor network applications. IET Commun. 14(5), 715–722 (2020)
9. Hoang, D.C., Yadav, P., Kumar, R., Panda, S.K.: Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks. IEEE Trans. Ind. Inf. 10(1), 774–783 (2014)
10. Xu, L., Collier, R., O'Hare, G.M.P.: A survey of clustering techniques in WSNs and consideration of the challenges of applying such to 5G IoT scenarios. IEEE Internet Thing 4(5), 1229–1249 (2017)
11. Handy, M.J., Haase, M., Timmermann, D.: Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In: *Proceedings of the 4th International Workshop on Mobile and Wireless Communications Network*. Stockholm, Sweden, pp. 368–372 (2002)
12. Lindsey, S., Raghavendra, C.S.: PEGASIS: Power-efficient gathering in sensor information systems. In: *Proceedings of the IEEE Aerospace Conference*. Big Sky, MT, pp. 1125–1130 (2002)

13. Manjeshwar, A., Agrawal, D.: TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In: *Proceedings of the 15th International Parallel and Distributed Processing Symposium* (IPDPS 2001). California, pp. 2009–2015 (2001)
14. Lin, D., Wang, Q.: A game theory based energy efficient clustering routing protocol for WSNs. Wireless Networks 23(4), 1101–1111 (2017)
15. Lin, D., Wang, Q., Lin, D., et al.: An energy-efficient clustering routing protocol based on evolutionary game theory in wireless sensor networks. Int. J. Distrib. Sens. Networks 2015(1), 1–11 (2015)
16. Gupta, I., Riordan, D., Sampalli, S.: Cluster-head election using fuzzy logic for wireless sensor networks. In: *Proceedings of the 3rd Annual Communication Networks and Services Research Conference* (CNSR 2005). Halifax, NS, Canada, pp. 255–260 (2005)
17. Lee, J., Cheng, W.: Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication. IEEE Sens. J. 12(8), 2891–2897 (2012)
18. Wang, Q., Lin, D., Yang, P., et al.: A fuzzy-logic based energy-efficient clustering algorithm for the wireless sensor networks. In: *Proceedings of the 2018 26th International Conference on Software, Telecommunications and Computer Networks* (SoftCOM 2018). Split, Croatia, pp. 1–6 (2018)
19. Latiff, N.M.A., Tsimenidis, C.C., Sharif, B.S.: Energy-aware clustering for wireless sensor networks using particle swarm optimization. In: *Proceedings of the 2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*. Athens, Greece, pp. 1–5 (2007)
20. Osamy, W., El-Sawy, A.A., Salim, A.: CSOCA: Chicken swarm optimization based clustering algorithm for wireless sensor networks. IEEE Access 8(1), 60676–60688 (2020)
21. Kaur, T., Kumar, D.: MACO-QCR: Multi-objective ACO-based QoS-aware cross-layer routing protocols in WSN. IEEE Sens. J. 21(5), 6775–6783 (2021)
22. Cai, X., Wang, P., Du, L., Cui, Z., Zhang, W., Chen, J.: Multi-objective three-dimensional dv-hop localization algorithm with NSGA-II. IEEE Sens. J. 19(21), 10003–10015 (2019)
23. Wang, P., Huang, J., Cui, Z., Xie, L., Chen, J.: A gaussian error correction multiobjective positioning model with NSGA-II. Concurrecy Comput. Pract. Exp. 32(5), 1–16 (2020)
24. Zheng, J., Wang, P., Li, C.: Distributed data aggregation using slepian-wolf coding in cluster-based wireless sensor networks. IEEE Trans. Veh. Technol. 59(5), 2564–2574 (2010)
25. Abasikeleş-Turgut, İ.: DiCDU: Distributed clustering with decreased uncovered nodes for WSNs. IET Commun. 14(6), 974–981 (2020)
26. Singh, B., Lobiyal, D.K.: An energy-efficient adaptive clustering algorithm with load balancing for wireless sensor network. Int. J. Sens. Networks 12(1), 37–52 (2012)
27. Murata, T., Ishibuchi, H.: Performance evaluation of genetic algorithms for flowshop scheduling problems. In: *Proceedings of 1st IEEE Conference on Evolutionary Computation*. Orlando, FL, pp. 812–817 (1994)
28. Muruganathan, S., Ma, D., Bhasin, R., et al.: A centralized energy-efficient routing protocol for wireless sensor networks. IEEE Commun. Mag. 43(3), S8–13 (2005)
29. Behera, T.M., Mohapatra, S.K., Samal, U.C., Khan, M.S., Daneshmand, M., Gandomi, A.H.: Residual energy-based cluster-head selection in WSNs for IoT application. IEEE Internet Thing 6(3), 5132–5139 (2019)
30. Hu, Y., Niu, Y., Lam, J., Shu, Z.: An energy-efficient adaptive overlapping clustering method for dynamic continuous monitoring in WSNs. IEEE Sens. J. 17(3), 834–847 (2017)
31. Hoang, D.C., Yadav, P., Kumar, R., Panda, S.K.: Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks. IEEE Trans. Ind. Inf. 10(1), 774–783 (2014)
32. Zhang, Y., Xing, S., Zhu, Y., Yan, F., Shen, L.: RSS-based localization in WSNs using gaussian mixture model via semidefinite relaxation. IEEE Commun. Lett. 21(6), 1329–1332 (2017)
33. Lin, D., Gao, L., Min, W.: A social welfare theory-based energy-efficient cluster head election scheme for WSNs. IEEE Syst. J. 15(3), 4492–4502 (2021)
34. Nabeel, M., Dressler, F.: Experimental evaluation of receive diversity techniques in distributed sensor networks. Ad Hoc Networks 99(15), 1–10 (2020)
35. Lin, D., Wang, Q., Liu, J.: Research on energy-efficient inter-cluster routing algorithm based on non-cooperative game. J. Harbin Inst. Technol. 49(11), 95–100 (2017)
36. Perera, T.D.P., Jayakody, D.N.K., Sharma, S.K., Chatzinotas, S., Li, J.: Simultaneous wireless information and power transfer (SWIPT): Recent advances and future challenges. IEEE Commun. Surv. Tutorials 20(1), 264–302 (2018)
37. Lu, Y., Xiong, K., Fan, P., Ding, Z., Zhong, Z., Letaief, K.B.: Secrecy energy efficiency in multi-antenna SWIPT networks with dual-layer PS receivers. IEEE Trans. Wireless Commun. 19(6), 4290–4306 (2020)
38. Stirling, W.C., Frost, R.L.: Social utility functions-part II: Applications. IEEE Trans. Syst. Man Cybern. Part A Syst. Humans 35(4), 533–543 (2005)
39. Atkinson, A.B.: On the measurement of inequality. J. Econ. Theory 2(3), 244–263 (1970)

How to cite this article: Lin, D., Kong, L., Zhao, C., Gao, J., Ouyang, H., Yang, Z., Zhang, Z.: An energy-efficiency-adaptive clustering formation mechanism for the wireless sensor networks. IET Commun. 16, 255–265 (2022). https://doi.org/10.1049/cmu2.12343

APPENDIX A

Concept of social welfare

The concept of social welfare was proposed to evaluate the equality degree of human society quantitatively [38]. Consequently, the Akinson's Social Welfare Function was presented to evaluate the human quality quantitatively [24, 39]. Similar to the human society, all the sensor nodes cooperate with each other to perform a complicated task. Therefore the concept of social welfare is also suitable to deal with the relationship between each sensor node and the whole network. Obviously, it complies with the principle of WSNs that a more balanced energy distribution contributes to a longer network lifetime. Therefore, we adopt the Atkinson's social welfare function to promote energy equilibrium within each cluster.

APPENDIX B

Definition of energy efficiency welfare

Definition 4.2. Energy efficiency welfare (E^2W) is a concept which reflects the condition of both residual energy and energy equality concurrently. To be specific, it is defined as the product of EI_{ec} and the residual energy in this paper.

Based on the concept of EI_{ec} proposed in Section 3, the concept of E^2W can be mathematically defined as follows

$$E^2W(i) = E_{re}^i \cdot EI_{ec} = E_{re}^i \cdot \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{d_{i-CH}^2}{\overline{d^2}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} \tag{19}$$

where $E^2W(i)$ denotes the energy efficiency welfare and E_{re}^i the expected residual energy of sn_i when it is considered as the CH in the next round.

Definitions 4.1 and 4.2 provide an effective evaluation criterion for the process of CH election actually. In fact, both of the residual energy and the energy equality are taken into consideration concurrently. As a result, the node generating a higher energy efficiency welfare contributes to a higher energy efficiency. Therefore, the node can be elected as the optimal CH if it produces the largest value of E^2W.