IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 15 June 2022, accepted 28 June 2022, date of publication 6 July 2022, date of current version 11 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188258

RESEARCH ARTICLE

Affinity Propagation and Chaotic Lion Swarm Optimization Based Clustering for Wireless Sensor Networks

HU HUANG-SHUI[1], GUO YU-XIN[1], WANG CHU-HANG[2], AND GAO DONG[1]
[1]College of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China
[2]College of Computer Science and Technology, Changchun Normal University, Changchun 130026, China

Corresponding author: Wang Chu-Hang (526213804@qq.com)

This work was supported in part by the Science and Technology Development Project of Jilin Province under Grant 20200201009JC, Grant 20210201051GX, and Grant 20210203161SF; and in part by the Education Department Project of Jilin Province under Grant JJKH20220839KJ.

ABSTRACT By grouping nodes with similar attributes into clusters, the energy efficiency and lifespan of wireless sensor networks(WSNs) can be improved effectively. However, the number of clusters needs to be set in advance, and the optimal cluster heads (CHs) are difficult to determine, which will undoubtedly reduce the network's overall performance. Therefore, a new clustering method using Affinity Propagation and Chaotic Lion swarm Optimization is proposed in this paper to form optimal clusters, which is called APCLO. In APCLO, Affinity Propagation (AP) is used to construct a cluster topology, forming the initial clustering according to the remaining energies of the nodes and the similarity of the distances between nodes. Moreover, the initial CHs are also determined simultaneously by AP. In order to eliminate the outliers from the initial CHs, Chaotic Lion Optimization(CLO) is presented to find the best CHs, in which a fitness function is set according to residual energy and distance to the BS, and chaotic map is used to speed up the convergence of CLO. Simulation results show that the APCLO protocol is superior to the comparison protocols in terms of energy consumption, network throughput, convergence speed, and lifetime. For network lifespan, it increases by 20.1%, 11.2%, and 13.5% respectively.

INDEX TERMS Wireless sensor networks, affinity propagation, chaotic lion swarm optimization, cluster head selection, minimizing energy consumption.

I. INTRODUCTION

Along with the development of the Internet of Things, wireless sensor networks (WSNs), as one of its key technologies, have gradually penetrated all areas of human life, such as environmental monitoring and forecasting, logistics distribution, health care, and space exploration [1]. In WSNs, a large number of sensor nodes and one or more Base Station(BS) cooperate to complete the tasks such as information collection, processing, and transmission. Every node is equipped with restricted energy, storage, and processing capabilities. Thus, energy saving has been the most critical research focus to extend the network lifespan of WSNs, and clustering has been considered the most energy-saving method [2]. In cluster-based WSNs, clusters are formed by grouping all the sensors. A set of special nodes called CHs are selected from the clusters, which are responsible for timeslot allocation and data collection from cluster numbers (CMs), data aggregation, and forwarding. Moreover, a round is usually used to rotate the CHs in the clusters [3]. The total network energy consumption is reduced mainly by clustering because of the short distance communication between CMs and CHs, less data transmission between CHs and BS, periodic sleep of CMs, and role transfer among CHs and CMs. Therefore, a great many clustering protocols have been proposed to prolong the network lifespan during the last decades [4].

In a cluster, cluster rotation is a common technique for balancing energy dissipation and improving network lifespan.

The associate editor coordinating the review of this manuscript and approving it for publication was Hosam El-Ocla.

Each CH consumes more energy compared to a CM [5]. Thus the number of CHs and their selection play a critical role in extending the network lifespan. Determining the number of clusters has an essential impact on the network's longevity because fewer clusters in the network inevitably cause CMs to communicate with CHs over a longer distance. At the same time, too many clusters lead to long-distance communication between CHs and the BS. Most clustering methods use a fixed number of clusters during the clustering process [6], which can't adapt to the scale of the nodes, resulting in low energy efficiency. Therefore, AP is used to solve this problem. AP does not need to set any CHs and can adaptively make the existing data points as initial CHs. Compared with K-means and other clustering methods, it has a less square error of results [7]. However, the general AP algorithms usually form clusters according to the distance similarity matrix and often ignore the residual energy and other factors.

After determining the number of clusters, the most important thing is to select the best CHs. Traditional probability-based or weight-based methods to select CHs based on the random preset values or calculated weight values cannot solve the problem [8], [9]. Therefore, intelligence-based optimization methods are used to obtain the approximate solutions for CHs selection due to their local or global search capabilities such as particle swarm optimization, imperialist competitive algorithm, genetic algorithms [10], lion swarm optimization, and so on. When the appropriate CHs are selected, the remaining ordinary nodes join the clusters according to different parameters such as distance and remaining energy, forming a unified and energy-saving cluster.

As described above, the AP clustering algorithm can perform adaptive clustering according to the network without setting the number of clusters. In addition, the clustering method based on swarm intelligence has become the latest scheme to improve energy efficiency and prolong the network life because of its scalability, adaptability and global search ability. For example, in a dynamic and uncertain WSN, the lion optimization algorithm can obtain the best possible solution selected by CHs. Moreover, its low complexity is more suitable for WSNs than particle swarm optimization and other soft computing-based approaches. Therefore, this paper presents a hybrid Affinity Propagation (AP) and Chaotic Lion Optimization algorithm called APCLO to form optimal clusters. In APCLO, without the need to determine the cluster number in advance, AP and CLO are used to form energy-efficient and balanced clusters so as to maximize the network lifespan. The main contributions of this work are summarized as follows.

· AP with novel preference is used to form energy-efficient clusters, which makes nodes grouped together according to nodes' residual energy and the similarity of the distance between the nodes.

· Chaotic lion optimization algorithm is adopted to select the best CHs, which uses chaotic mapping to speed up the convergence and find the optimal CHs by defining a new fitness function to minimize the energy consumption of intra-cluster communication.

· Simulation experiments are performed to verify the effectiveness of APCLO in terms of energy consumption, network throughput and lifespan compared with the other protocols.

The remainder of this paper is organized as follows. The related works are discussed in Section 2, and the network model is described in Section 3. In Section 4, the proposed APCLO is introduced in detail. In Section 5, simulations are performed, and results are analyzed in sequence. Finally, the conclusion is made in Section 6.

II. RELATED WORKS

Since the pioneering clustering protocol low-energy adaptive clustering hierarchy (LEACH) [11] adopted clusters to organize the nodes with advantages such as less overhead by clustering and aggregation and long network lifetime by rotating CHs in rounds, numerous methods had been presented to improve the performance of the network from clustering and routing [12]. Generally, clustering schemes built cluster topology for routing protocols so as to find the best paths of data communication, which was the decisive factor in the network performance. Therefore, the following introduction only focuses on schemes for cluster formation.

The multi-level clustering protocol LEACH-SWDN algorithm proposed in 2012 added energy to the threshold design based on the LEACH protocol, which overcame the problem of reducing the threshold due to energy reduction [13]. However, there were still many parameters, such as the distance from nodes to the BS and the distance from nodes to the CHS. Therefore, *Suniti Dutt* proposed a CH constrained energy-saving routing protocol (CREEP) for heterogeneous WSNs in 2018 [14]. In order to improve the energy-saving effect, a distance component was introduced into the probability equation in CREEP protocol, which ensured that the probability of nodes far away from the BS becoming CHs was small.

In addition to some traditional algorithms improved based on LEACH protocol, some scholars introduced typical clustering methods such as K-means and C-means into WSN clustering protocol [15]. Although the above methods effectively solved some problems existing in LEACH, there would be the problem that the weight of various parameters can not be determined with the increasing variables. Thence, many scholars proposed using fuzzy control to solve the problem [16]. For example, *Sert et al.* proposed a two-tier distributed fuzzy-logic based protocol called TTDFP [17]. In the clustering phase, TTDFP used three fuzzy parameters: node connectivity, distance to BS, and remaining node energy. In the routing phase, TTDFP used FLS to extend the operational architecture of a known multihop routing method. *Mazinani et al.* proposed two FLS to select CHs in FMCR-CT proposed in 2019 [18]. Fls1 used parameter residual energy and node density as descriptors. Nodes with more energy and higher density were more likely to be selected as CHs. In addition, once the residual energy of any CHs selected by Fls1 was less than the threshold, Fls2 was triggered to select CHs. Although fuzzy control can constrain multiple parameters simultaneously, the optimal number of CHs was involved in the above protocols.

In 2021, *Panchal and Singh* proposed an energy-saving technology to select the optimal number of CH and grid heads named EOCGS [19]. EOCGS gave the expression of the optimal number of clusters, then proposed a new method to select CHs, in which the optimal number of clusters was calculated by a geometric method. AP did not need to specify the cluster number in advance as a new clustering algorithm. It only solved the problem by setting several parameters, which was better than the scheme based on geometry and soft computing. LEACH-AP proposed by *Sohn et al.* in 2016, used the negative energy consumption between nodes to define similarity so that nodes with larger values belong to the same cluster [20]. Additionally, LEACH-AP defined a preference expressed in terms of self-similarity. Although LEACH-AP can form an optimal cluster, the transmission distance within the cluster was long, and the energy efficiency of the network would be significantly reduced. *Wang et al.* proposed an affinity propagation-based self-adaptive (APSA) clustering method in 2019 [21]. The negative Euclidean distance was used to represent the similarity between two nodes so that the nodes close to each other became a cluster. Then the nodes with large energy were used to randomly replace the original CHs to continuously reduce the total distance from the member nodes to CHs. The clusters formed in APSA were more uniform than those formed in LEACH-AP, but the fixed preference made it challenging to select the best CHs.

In recent years, clustering methods based on swarm intelligence have been used to obtain approximate solutions of CH selection because they have local or global search ability [22]. For example, the multi-objective fractional particle swarm optimization (MOFPL) algorithm for energy-aware routing proposed by *Bhardwaj and Kumar* in 2019 defined a multi-objective fitness function based on energy, delay, traffic rate, distance, and clustering density [23]. Then, the lion swarm algorithm was used to continuously optimize the selection of CHs until the CH satisfying the maximum value of fitness function is found. In 2019, *Karthick* and *Palanisamy* proposed the krill herd (KH) optimization algorithm to optimize the selection of CHs [24]. The KH algorithm had the advantage of solving structural engineering problems simply. The protocol first randomly selected a set of CHs, and then set a membership function based on the average distance from the nodes to CH and the energy of CH. Although the protocol reduced energy consumption by continuously optimizing function values through KH, all CHs needed to be optimized for each iteration so that some suitable nodes for CH could not be selected. In addition to these algorithms, a music-based metaheuristic optimization algorithm called Harmonic Sound Search (HSA) has also been applied to the field of WSN. In 2020, an efficient search algorithm using the dynamic capabilities of fuzzy logic and HSA was proposed, which optimizes the distance between clusters, the transmission distance of CH, and the energy balance of CH to achieve the purpose of prolonging network life [25]. Thus, the intelligent optimization algorithm can also be combined with other methods. The SF-MPSO clustering protocol based on fuzzy control proposed by *Lipari et al.* in 2021 used a particle swarm optimization (PSO) algorithm to generate fuzzy rules [26]. However, using this algorithm to optimize CHs would have problems such as low convergence speed or falling into local optimal solutions. In order to address such problems, *Subramanian et al.* proposed an idea of combining gray wolf optimization algorithm and crow search optimization (CSO) in 2020 [27]. Gray wolf had the potential to prevent the stagnation of optimal local detection to a great extent, but its search depth was not enough. CSO search algorithm can prevent falling into a locally optimal solution. Similarly, *Alghamdi* proposed a CH election model based on hybrid dragonflies and firefly optimization algorithm (FPU-DA), which considered four significant factors: energy consumption, latency, distance, and security [28]. FPU-DA combined two algorithms to optimize the weight coefficients of membership functions and improve the convergence rate. The safe and energy-efficient clustering algorithm (eeTMFO/GA) proposed by *Sharma et al.* used both moth flame optimization and genetic algorithm [29]. It used MFO to optimize a randomly generated initial population and cross-update the population via GA. It can be seen that the combination of the two algorithms can solve the above problems to some extent. Therefore, this paper presented an idea of applying a chaotic map algorithm to the lion swarm algorithm, which used tent mapping to traverse uniformity to jump out of the optimal local solution and improve its convergence rate. Finally, combined with the paper mentioned above, a clustering algorithm APCLO, which combined Affinity Propagation and Chaos Lion Optimization, is proposed. In addition to the above brief introduction, we have listed the key strengths and weaknesses of some of the above protocols in Table 1 [30], [31].

III. NETWORK MODEL

In clustered WSNs, a certain number of nodes with limited resources are randomly scattered in the target sensing area, and nodes in the vicinity are grouped into a cluster. All nodes belong to a corresponding cluster, and a CH is selected to manage the cluster in each cluster. For any CMs, it communicates with their CH only. Meanwhile, the CHs communicate with the BS [32]. Then, energy is inevitably consumed during these communications. The widely used first-order radio model like in [21, 24, 25] is applied in this paper. When a node i sends $l-$ bits data to node j, its amount of energy consumption can be obtained as equation (1).

$$E_T(i,j) = \begin{cases} l \times E_{elec} + l \times E_{fs} \times d_{ij}^2, & d_{ij} \le d_0 \\ l \times E_{elec} + l \times E_{mp} \times d_{ij}^4, & d_{ij} > d_0 \end{cases} \quad (1)$$

where E_{elec} denotes the energy consumed to transmit or receive 1-bit data, E_{fs} and E_{mp} indicate the amplifier coefficients of free space and multi-path fading respectively, d_0 means the threshold distance given by $d_0 = \sqrt{E_{fs}/E_{mp}}$.

TABLE 1. Comparison of the proposed protocols.

Protocol	Strength	Weakness
LEACH	Rotate CHs for clustering	Lack of consideration of energy and distance
LEACH-SWDN	Increase energy in threshold design	Ignore the distance from CH to BS
CREEP	Introduce distance component	Ignore the factor of node degree
FMCR-CT	Adjust weights using FLS	The number of CHs cannot be obtained
TTDFP	The competition radius and CHS of nodes are determined through FLS	Parameter fuzzy processing will reduce the accuracy
EOCGS	Pre calculated quantity of CHs	Uneven distribution of CHS
LEACH-AP	Adaptive clustering using AP	Poor network scalability
APSA	Optimize distance using K-Means	The preference in AP algorithm is fixed
MOFPL	Energy, delay, traffic, distance	Ignore the weight of each parameter
KH	Optimize CHs with KH based on distance and energy	Low convergence rate
HSA	Using HSA to optimize transmission distance and energy	CHs load imbalance
SF-MPSO	Combine FLS and PSO to optimize CHs	Slow convergence or falling into local optimal solution
HGWCSOA-OCHS	Prevent local optimal detection stagnation	The search depth is not enough.
FPU-DA	Optimize the weight coefficients and improve the convergence rate	Lack of consideration of energy consumption balance
eeTMFO/GA	Improve the convergence speed through GA	CHs load imbalance

At the same time, the amount of energy for node i consumed to receive $l - bits$ data from node j can be calculated by equation (2)

$$E_R = l \times E_{elec} \tag{2}$$

In addition, the amount of energy consumption for aggregating $l - bits$ data is given by equation (3)

$$E_{DA} = l \times E_{da0} \tag{3}$$

where E_{da0} denotes the energy consumption for 1-bit data fusion. Our main goal is to reduce the energy consumption and uniformly distribute clusters so as to extend the network lifespan. To this end, the following assumptions are made for the presented network.

- The WSN with n nodes identified by ID is considered to be homogeneous after being deployed.
- All nodes are equipped with same initial energy, and can act as CHs or CMs. Moreover, they can sense their surroundings and send the sensed data to the BS or another node.
- The BS is also static and has unlimited energy and other resources.
- Each node can adjust its transmission power according to the distance of the receiver node.

Besides, some used symbols are listed as follows.

- N_i denotes the set of neighbors of node i and $|N_i|$ is the number of N_i.
- $Eres_i$ denote the initial energy the residual energy of node i respectively.
- d_{ij} is the distance between nodes i and j, and $d_{\max}$ is the maximum communication range for each node.
- Ld_i denotes the load of node i.
- rEB_i denotes the ratio of residual energy to each load for node i, which is expressed as $rEB_i = Eres_i/Ld_i$.

IV. THE PROPOSED CH SELECTION SCHEME

In this section, the AP algorithm is firstly adopted to determine the optimal number of clusters and form the initial clusters. Then, the chaos lion optimization algorithm is utilized to find the optimal CHs by iterations based on the initial clusters. Compared to the traditional AP application, the proposed scheme can select more reasonable CHs. Accordingly, it can improve the network energy efficiency and prolong the network lifespan.

A. FORMING INITIAL CLUSTERS BASED ON AP

After the WSN deployment is completed, the BS broadcasts a message to all nodes, and the nodes calculate the distance from BS according to the RSSI according to the received message [33]. Each node then records the combined remaining energy information and sends it to the BS. Accordingly, the BS obtains the information of all nodes for the AP to perform clustering to find the optimal number of clusters and the position of the initial cluster center. It is necessary to minimize the distance between CMs and their CHs to reduce the energy consumption of intra-cluster communication. Therefore, using the negative absolute value of the distance difference between nodes i and j to calculate their similarity is expressed as equation (4).

$$s(i,j) = -\left|S(i_1, i_2) - S(j_1, j_2)\right| i, j \in [1, n], i \in N_j, i \neq j \tag{4}$$

where $s(i_1, i_2)$ is the coordinate position of node i in the two-dimensional network. Moreover, the value of $s(i,j)$ sets to minus infinity when node i can't directly communicate with node k. In addition, the preference $s(j,j)$ indicates that node j will be selected as CH is given by equation (5)

$$s(j,j) = \frac{M}{1+\alpha} + \beta \tag{5}$$

where:

· α represents the normalized degree of the distance of node j from the BS, as shown in equation (6)

$$\alpha = \frac{DiBS_j - DiBS_{\min}}{DiBS_{\max} - DiBS_{\min}} \tag{6}$$

$DiSB_j$ represents the distance from the node to the BS, $DiSB_{\min}$ represents the minimum value of the distance to the BS among the neighbor nodes of node j, and $DiSB_{\max}$ is the maximum value among them

· β denotes the normalized ratio of node $j's$ rEB to the average rEB of its neighbors which can be calculated by equation (7).

$$\beta = \frac{EBavg_j - EBavg_{j\,\min}}{EBavg_{j_{\max}} - EBavg_{j\,\min}} \tag{7}$$

$EBavg_j = \frac{rEB_j}{\sum_{i \in N_j} rEB_i / |N_j|}$ $EBavg_{j\,\min}$ and $Eravg_{j\,\max}$ are the minimum and maximum of $EBavg_i = \frac{rEB_i}{\sum_{i \in N_j} rEB_i / |N_j|}$. Nodes with more residual energy and lower load and average similarity of neighbors have greater chance to be selected as CHs. Furthermore, residual energy as well as distance is also used to update responsibility *r(i,j)* and availability *a(i,j)* using equation (8) and (9) respectively. The former means the degree to node *j* selected as the CH of node *i,* and the latter reflects the appropriate degree of node *i* to select j as its CH.

$$r(i,j) = \frac{rEB_j}{d_{ij}} s(i,j) - \max_{\substack{j \neq j \\ j \in N_i}} \left(a\left(i,j'\right) + s\left(i,j'\right) \right) \tag{8}$$

$$a(i,j) = \begin{cases} \sum\limits_{\substack{i \neq j \\ i' \in N_k}} \max \left\{ 0, \frac{rEB_{i'}}{d_{i'j}} * r\left(i',j\right) \right\}, & if\, i = j \\ \min \left\{ 0, r(j,j) + \sum\limits_{\substack{i \notin \{i,j\} \\ i \in j}} \max \left(0, \frac{rEB_{i'}}{d_{i'j}} * r\left(i',j\right) \right) \right\}, & if\, i \neq j \end{cases} \tag{9}$$

Nodes with more energy and smaller average distance to neighbors are more likely to be selected as CHs. Moreover, the initial value of *a(i,j)* is set to zero. By using the corresponding values in the last iteration, the updating process continues until the termination condition is satisfied, that is, reaching the preset number of iterations or no more improvements in the solution. At present, the best solution is obtained, and k nodes meeting with $r(j,j) + a(j,j) > 0, j \in [1, n]$ are selected as CHs. The other nodes determines their relevant CHs according to the similarity values.

B. FINDING THE OPTIMAL CHS BASED ON LOA

Generally, there may be outliers for the initial CHs determined by AP, which means that the selected CHs are non-optimal. So the Loin Swarm Optimization is used to find the optimal CHs based on the initial clusters [34]. For any CH, the position of its member nodes can be denoted by a vector $X_i = (x_1, x_2, \cdots x_j)$, $(j \in [1, nM_i])$, nM_i is the number of nodes in the CH_i cluster, and k is the size of the search space. During the process of optimization, for minimization of the energy consumption of intra-cluster communication, the energy consumption of intra-cluster communication as well as the ratio of the residual energy of the CHs and their distance to the BS are considered for definition of the fitness function. The energy consumption of intra-cluster communication of CH_i is mainly composed of three parts: the energy of member nodes transmitting data to CHs, and the energy of CHs receiving and fusing data. The total energy in the cluster can be expressed by equation (10).

$$inEss_i = \sum_{j=1}^{nM_i} E_T\left(x_j, CH_i\right) + nM_i \times (E_T + E_{DA}) \tag{10}$$

In the above formula, $E_T = (x_j, CH_i)$ is the energy consumed by the member node to transmit data to its CH_i, and E_R and E_{DA} are the energy consumed by the ch to receive and fuse data respectively. Then, the normalized value of energy consumption of intra-cluster communication for CH_i can be obtained by equation(11).

$$inEs(i) = \frac{inEs_i - inEs_{\min}}{inEs_{\max} - inEs_{\min}} \tag{11}$$

where $inEs_{\min}, inEs_{\max}$ are the minimum and maximum energy consumed by each cluster in the network respectively. In addition, the ratio of the CH i' energy to its distance to the BS is used as another parameter,, which is shown as equation (12).

$$distE_i = \frac{Eres_i}{DiSB_i} \tag{12}$$

Also, the normalized value of $distE_i$ can be obtained by equation (13).

$$distE(i) = \frac{dist_i - distE_{\min}}{distE_{\max} - dist_{\min}} \tag{13}$$

where $distE_{\min}, distE_{\max}$ are the minimum and maximum of $distE_i, i \in [1, nM_i]$ respectively. Accordingly, the fitness function can be expressed as equation (14).

$$Fitness(i) = \frac{1}{1 + inEs(i)} + distE(i) \tag{14}$$

Each CH (Lion King) moves within its clustering range by equation (15), and compares the fitness value of each iteration to find the best position. Then the node closest to this position is selected as the new CH.

$$x_i^{k+1} = g^k \left(1 + \gamma \left\| p_i^k - g^k \right\| \right) \tag{15}$$

where g^k denotes the optimal position in the *kth* iteration, p_i^k is the best position during the historical iterations for lion i, and γ is a coefficient following normal distribution.

In this way, although a better CH can be found, the convergence speed is very slow and it is easy to fall into a locally optimal solution. In order to improve the convergence of the lion group algorithm, APCLO combines chaotic map in the iteration of LSO, which is an effective local search algorithm. In this paper, the method of Tent mapping is used for optimization iteration. After the AP completed the election of the original lion kings (CHs), the obtained optimal individual $CHx_k = (ch1, ch2, \cdots ch_k)$ is searched by Tent chaos[35]. Assume that the current search space range is $[u_k, l_k]$. $[u_k, l_k]$ is the maximum distance and the minimum distance between nodes in the cluster where the CH is located. The main steps of Tent map can be described as follows:

```
APCLO Clustering Protocol Pseudocode
Input: the coordinate set of N sensor nodes {X1, X2, X3 Λ Xn}
Initialize APCLO
rn=1
rounds=max (the number of rounds the network runs)
While rn <rounds
    rn=rn+1
        For i=1:n
            For j=1:n
                Forming clustering by AP algorithm
                Calculate CH={ch1, ch2,...... chk}
            END
    END
      For   CHk=1:k
          Updating CH by CLO
          Calculate new CH={ch1, ch2, ......, chk}
Until the CH remains unchanged for T rounds or reached max iterations
    End
End
```

FIGURE 1. The pseudo code diagram of APCLO.

Step 1: Transform the CHx_k, $CHz_k(t) = \frac{CHx_k - u_k}{l_k - u_k}$t $= 1, 2, \ldots, t_{\max}$, where $t_{\max}$ represents the maximum number of iterations of Tent map, and map it to the interval;

Step 2: Transform the chaotic variables into the interval range and generate new individuals, $CHy_k = CHx_k + \frac{u_k - l_k}{2} \times (2CHz_k(t) - 1)$. The CH in the next round are the node closest to CHy_k.

Step 3: Taking the minimum optimization as an example, if $Fitness(CHx_k) < Fitness(CHy_k)$ then $CHx_k = CHy_k$.

Step 4: If the optimal value is not updated in consecutive T rounds, or after reaching the maximum number of chaotic iterations, end the chaotic map. Otherwise, go to step 2.

The detailed flow diagram of APCLO is illustrated in Fig. 2. Once the CHs are selected, the BS broadcasts a message including the CHs and their corresponding CMs to all the nodes, and the nodes complete data transmission according to their roles. That is to say, each CM in the clusters sends the data attached with residual energy only to its CH, and the CH aggregates the data collected from its CMs for forwarding.

C. ROTATING THE CHS ON DEMAND

Periodic clustering in rounds is usually adopted to maintain the clusters for most clustering algorithms, which increases communication overhead with the amount of control information exchanges. However, this paper used on-demand clustering to rotate CHs properly. The CHs are selected and clustered by the set AP algorithm first, and then the CHs are updated in the cluster by running the CLO algorithm. Only local clustering can occur in the cluster. The detailed process of on-demand clustering is described as Fig.1.

D. TIME COMPLEXITY ANALYSIS

The APCLO algorithm is mainly composed of the AP algorithm and the CLO algorithm, so its time complexity can be expressed as O (APCLO)=O(time complexity of AP+time complexity of CLO). AP is composed of four

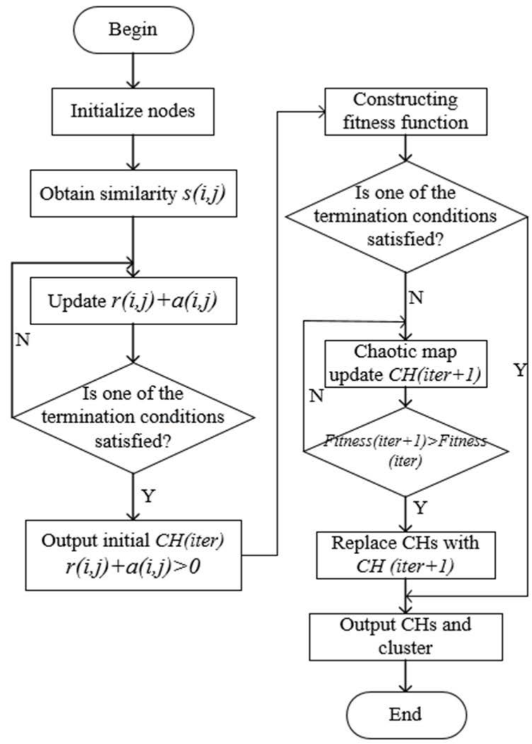


FIGURE 2. The flow diagram of APCLO.

parts: initialization, nodes similarity calculation,continuous and alternate updating of the Responsibility matrix and Availability matrix, as shown in formula (1), and judging the CHs through *r(j, j)+a(j, j)*. The time complexity of the initialization process is O (n^2), O ($n^2/2$) when calculating the node similarity, O ($n^2 * logn$) when updating *r(i,j)* and *a(i,j)*, and O (n) when judging the cluster center. Therefore, the time complexity of AP clustering process is $O(n^2 + n^2/2 + n^2 * logn + n)$. Once the iteration satisfies the termination condition, the optimal solution is found through bubble sorting, and the time complexity in the worst case is $O(n * n * logn)$. Then CLO is to iterate over the CHs generated by the AP. In the worst case, the CLO algorithm is executed for all CHs, and the time complexity is $O(k^2)$. k is the number of CHs selected by the AP algorithm. Moreover, k is less than n. So the time complexity of APCLO is $O(n^2.logn)$.

V. PERFORMANCE EVALUATION

This section will conduct the simulation experiments of APCLO in MATLAB 2019a. The experiments are carried out in the network with a sensing area of 200*200m. Because the location of the BS and the number of nodes will affect the experimental results, we set 100 and 200 nodes to be randomly distributed in the network, respectively, and the BS coordinates in each case are (100, 100) and (0, 0) in turn. In all scenarios, the energy of the sensor node battery is set to 1J. Other parameters and their values used in the simulation are shown in Table 2.

TABLE 2. Parameter settings.

Parameters	Values
Number of nodes	(100, 200)
Initial energy	1J
E_{elec}	50 (nJ/bit)
E_{pDb}	5 (nJ/bit)
E_{fs}	10 (pJ/bit/m^2)
E_{mp}	0.0013 ((pJ/bit/m^4)
d_0	87.7m
Data packet size	4000bits
Control packet size	200bits
BS Location	(x=100,y=100);(x=0,y=0)
Area	(200m*200m)

Four indicators, including the number of surviving nodes, total energy consumption, network throughput, and convergence speed, are used to test the performance of APCLO. It is compared with APSA, KH and HSA clustering algorithms proposed in [21], [24] and [25]. APCLO, APSA, KH and HSA protocols employ different algorithms to select and optimize CHs, which leads to differences in energy consumption and network lifespan. APSA adopts AP for clustering, whose preference is a fixed value, may lead to outliers in the selected CHs. Although K-medoids can reduce part of the error, it still ignores the consideration of the distance from CHs to BS. The KH algorithm updates all CHs simultaneously in each iteration, which will miss locally optimal CHs. As for HSA optimization algorithm, it can optimize the transmission distance but ignores the consideration of cluster head load balancing. APCLO considers energy and distance factors in defining AP preference and then continuously optimizes through CLO to find more suitable CHs. The simulation results below demonstrate the superiority of APCLO.

A. NETWORK LIFESPAN

In this experiment, the network lifespan is expressed by the number of remaining nodes in each round. Table 3 shows the simulation results of the first node death round (FND), half node death round (HND), and 80% node death round (END) for different protocols. In addition, the simulation results of the number of surviving nodes in each round are shown in Figure 3.

In these four protocols, KH is clustered in a fixed number of CHs. HSA load is unbalanced, which will lead to overload and premature death of some CHs in a network with many nodes randomly distributed. APSA and APCLO use the AP algorithm to determine the number of CHs and complete the initial clustering. However, APSA mainly focuses on the distance factor, ignoring the residual energy and loading of CHs. The preference setting of the AP algorithm improved by APCLO can be changed with the energy of the node, and the CHs generated by the initial clustering are optimized by using CLO.

According to the data in Table 2, in the four simulation scenarios, the first node of AOCLO dies in the 181,230,36 and 12 rounds, which is 63.7% higher than that of APSA protocol, 11.8% higher than that of the KH protocol, and 13.0% higher than that of HSA protocol on average. Half the number of APCLO nodes died in 1098, 884, 304 and 272 rounds respectively, which increased by 17.2%, 8.7%, 15.5% and 19.4% respectively compared with APSA, increased by 6.1%, 7.2%, 22.6% and 14.7% respectively compared with KH, and increased by 11.6%, 13.6%, 7.5% and 20.5% respectively compared with HSA. The number of rounds of 80% node death in APCLO protocol is 1.4% higher than that in APSA protocol, and 9.1% higher than that in KH protocol on average. Based on the above experimental data, the network lifetime of APCLO protocol proposed in this paper is extended by 20.1% on average compared with APSA protocol, and increased by 11.2% on average compared with KH protocol, and 13.5% longer than that of HSA protocol. At the same time, from the curve of the quantity of surviving nodes in the network in Figure3, it can be seen that APCLO has solved some problems in other protocols to a certain extent in prolonging the network lifespan and balancing the energy consumption of the network.

B. NETWORK ENERGY CONSUMPTION

The lifespan of a network is usually inversely proportional to the energy spending by the nodes. The smaller the energy spending, the smaller the transmission distance used when the network runs. The change in total network energy consumption is shown in Figure 4.

Seen from Figure 4, as the CH rotation frequency in the network increases, the network energy consumption keeps rising. The curves of network energy consumption for the KH and APSA algorithms remain above the APCLO protocol. This is because APSA and KH produce the large amount of energy consumed by the CHs to send data to the BS. However, APCLO selects nodes with smaller distances to the BS as CHs, thus reducing energy. It can also be concluded from the experimental data that APSA consumes half of its energy in rounds 523, 458, 122, and 144, KH in 519, 443, 155, and 123 rounds, and HSA in 492, 379, 125, and 99 rounds. In APLCO, the network energy consumption reaches 50% when CH rotations are 559, 484, 167, and 159, respectively. It can be drawn from these data, that in terms of reducing energy consumption, compared with APSA, KH and HSA, APCLO has increased by 11.4%, 12.2% and 24.1% on average, respectively.

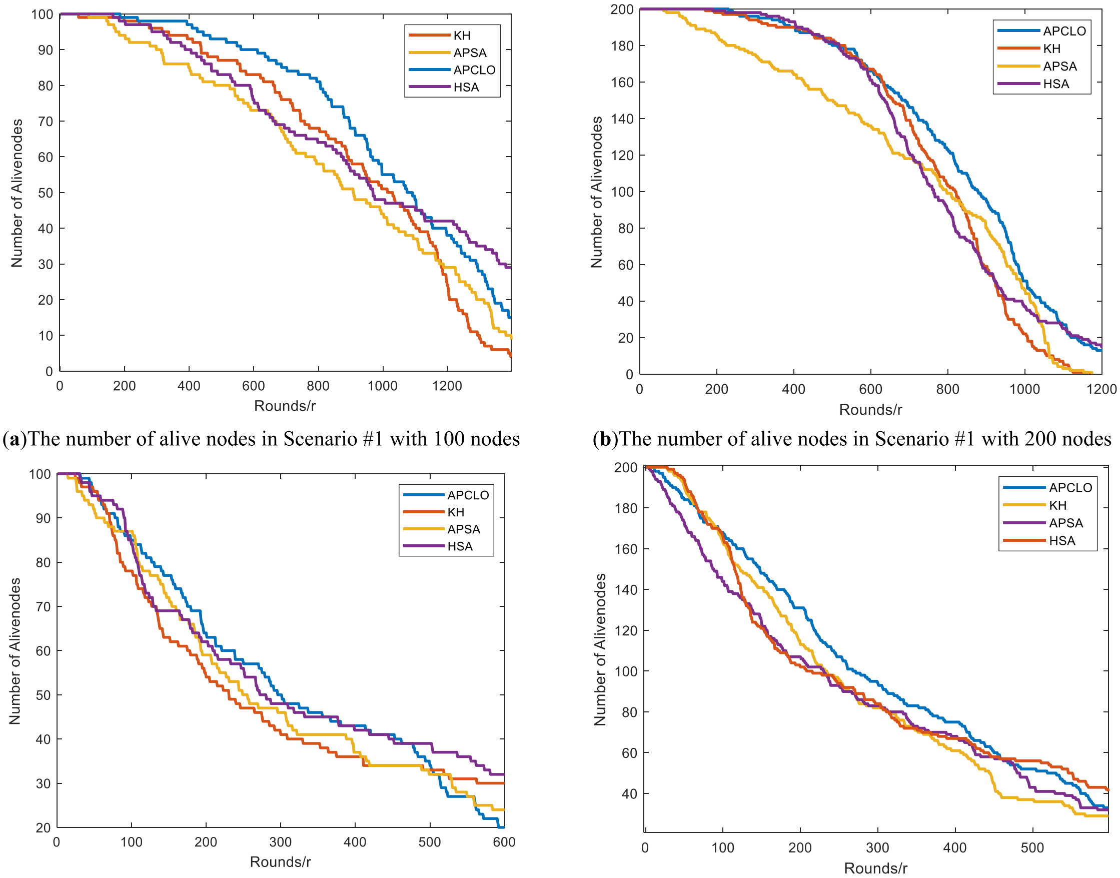
C. NETWORK THROUGHPUT

Network throughput indicates the number of messages transmitted by nodes to the BS. A higher amount of data transmission indicates a higher utilization rate of energy in the network. Figure 5 shows the results in the network throughput.

It can be seen from Figure 5 that for all protocols, their network throughput increases continuously with the number

TABLE 3. FND,HND and LND in different scenarios.

Items Protocols	Case	Status	APSA	KH	HSA	APCLO
Scenario #1	Case 1: 100 nodes	FND	58	89	164	181
		HND	909	1031	970	1098
		END	1302	1216	1526	1346
	Case 2: 200 nodes	FND	61	182	193	230
		HND	807	820	763	884
		END	1015	945	993	1046
Scenario #2	Case 1: 100 nodes	FND	16	25	31	36
		HND	257	235	281	304
		END	600	577	621	566
	Case 2: 200 nodes	FND	5	27	29	12
		HND	219	232	216	272
		END	535	458	602	565



(a)The number of alive nodes in Scenario #1 with 100 nodes

(b)The number of alive nodes in Scenario #1 with 200 nodes

(c)The number of alive nodes in Scenario #2 with 100 nod

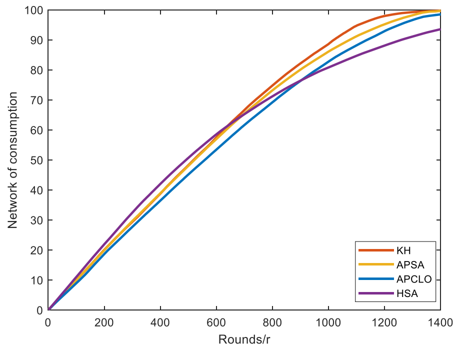
(d)The number of alive nodes in Scenario #2 with 200 nodes

FIGURE 3. Comparison of the network lifespan versus rounds.

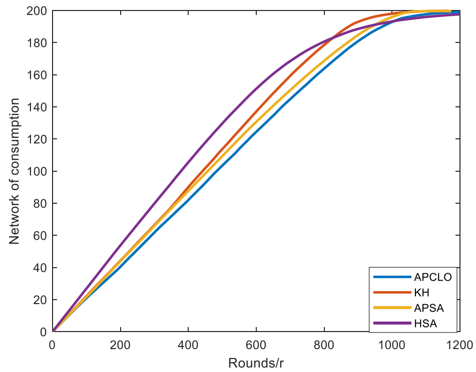
of rounds. In different protocols, the number of surviving nodes left in each round is different, so the amount of data transmitted will be various. APCLO achieved the highest total network throughput in the network throughput compared with HSA, KH and APSA. APCLO is 5.3%, 6.9%, 0.5%, and 6.1% higher than HSA, and 3.8%, 5.7%, 4.2%, and 10.2% higher than KH. Furthermore, APCLO is 16.9%, 16.0%, 12.9%, and 15.2% higher than APSA. The consequence can explain that the APCLO protocol not only effectively saves energy consumption but also ensures the total amount of data transferred.

D. CONVERGENCE SPEED

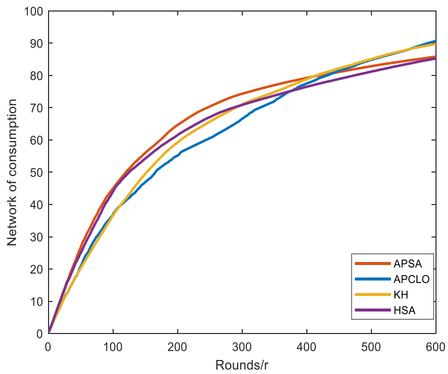
Convergence rate refers to the number of iterations of the protocol from initial clustering to final CH determination.



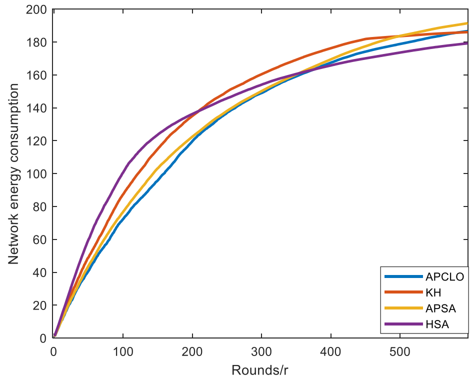
(**a**)Energy consumption in Scenario #1 with 100 nodes



(**b**) Energy consumption in Scenario #1 with 200 nodes

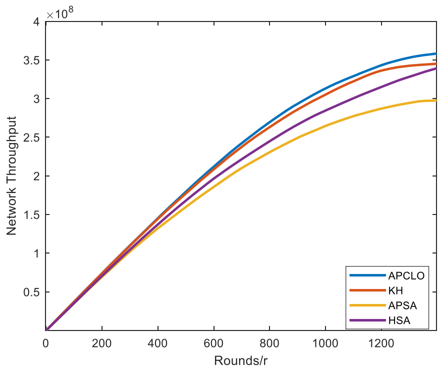


(**c**)Energy consumption in Scenario #2 with 100 nodes

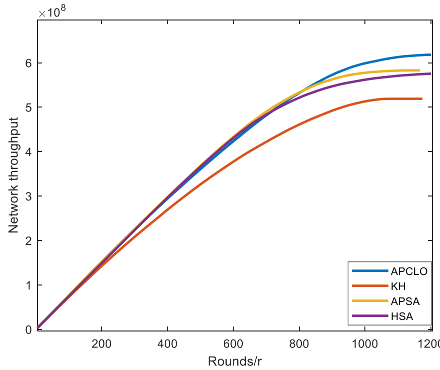


(**d**)Energy consumption in Scenario #2 with 200 nodes

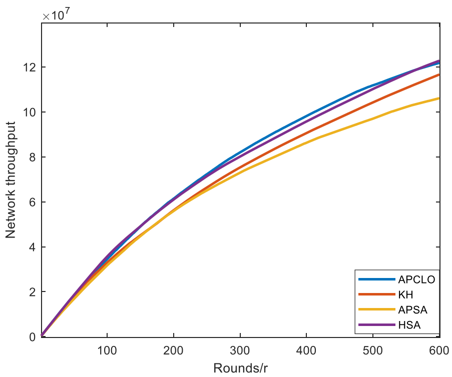
FIGURE 4. Comparison of the total energy consumption versus rounds.



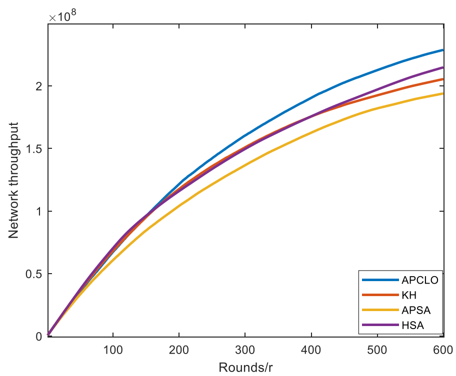
(**a**)The network throughput in Scenario #1 with 100 nodes



(**b**)The network throughput in Scenario #1 with 200 node

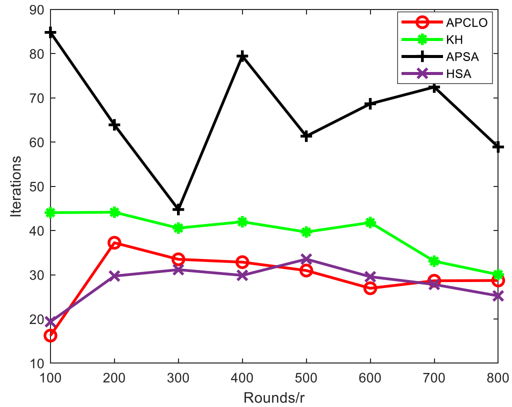


(**c**)The network throughput in Scenario #2 with 100 nodes

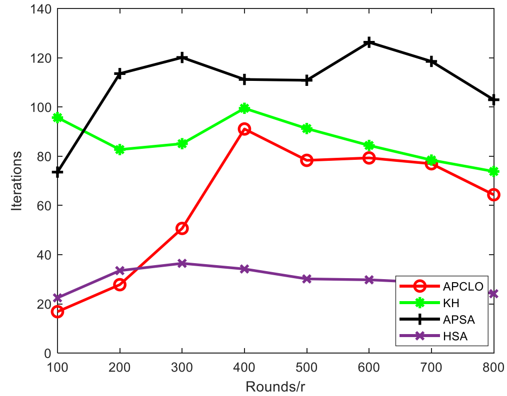


(**d**)The network throughput in Scenario #2 with 200 nodes

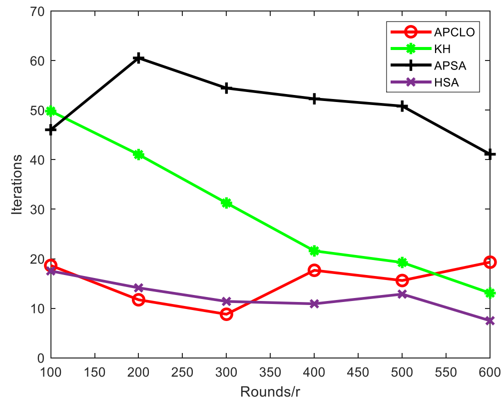
FIGURE 5. Comparison of the network throughput versus rounds.



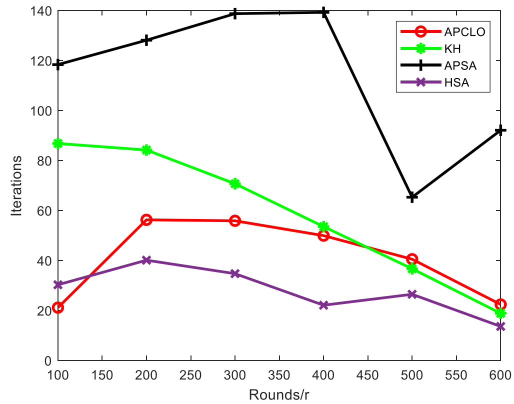
(**a**)The convergence speed in Scenario #1 with 100 nodes



(**b**)The convergence speed in Scenario #1 with 200 node



(**c**)The convergence speed in Scenario #2 with 100 nodes



(**d**)The convergence speed in Scenario #2 with 200 nodes

FIGURE 6. Comparison of convergence speed versus rounds.

In this paper, *Iter* is used to represent the algorithm's convergence rate. The convergence speed of the APCLO protocol *Iter*(APCLO)=(the number of initial clustering iterations performed by AP + the number of iterations required by CLO to optimize CH). By analogy, the convergence rates of the APSA,KH and HSA protocols are *Iter*(APSA)=*Iter*(AP)+*Iter*(K), *Iter*(KH) and *Iter*(HSA). The simulation results of the convergence rates of the four protocols are shown in Figure 6.

Since APSA adopts the stochastic principle to optimize the CHs derived by the AP algorithm, the convergence speed is slow. In the four scenarios of the experiment, the average convergence speed of APSA is 66.78, 109.7, 50.87, and 113.7 times, respectively. KH needs to update the entire set of CHs in each iteration during the optimization process, thus reducing its convergence speed. The iterations of KH are 39.42, 86.38, 29.34, and 58.54, respectively. Before HSA algorithm is implemented, the network has been processed by fuzzy logic system, so its convergence speed is relatively fast, and its convergence speed is 28.3, 29.9, 12.3 and 27.9. APCLO optimizes CH in a local range through the chaotic lion algorithm, improving convergence speed. The convergence rates of APCLO are 29.39, 60.67, 15.29, 41.09. From the experimental data, it can be concluded that the convergence speed of APCLO is 30.72% and 58.61% lower than that of KH and APSA, respectively.

TABLE 4. FND, HND and LND in the area of 1000m*1000m with different nodes.

Nodes / Protocols		APCLO	APSA	KH	HSA
600Nodes	FND	1	1	1	1
	HND	19	12	18	16
	LND	117	102	59	126
800Nodes	FND	1	1	1	1
	HND	21	11	20	17
	LND	140	112	117	102
1000Nodes	FND	1	1	1	1
	HND	19	12	19	16
	LND	127	120	106	135

E. NETWORK SCALABILITY

In recent years, the area and node usage of wireless sensor networks have increased exponentially, so clustering algorithms often need to be applied to larger network scenarios to prove the scalability of APCLO protocol. Therefore, in addition to the above experiments, we tested the network lifetime of APCLO when the network area is 1000*1000m. The experimental results are shown in Table 4.

It can be seen from Table 4 that in large-scale networks, the network life of nodes will be dramatically shortened because the distance between clusters will also increase, which will lead to the exponential growth in energy consumption. Because the protocol CHS proposed in this paper directly transmits the information of the member nodes to the BS, in this case, the CHS far away from the BS dies quickly, so FND usually occurs in the first round. However, it can be seen from Table 4 that even in large-scale networks, the APCLO protocol proposed in this paper has slight advantages over HSA, KH, and APSA protocols, and has better scalability.

VI. CONCLUSION

This paper proposed a clustering protocol APCLO based on affinity propagation and chaotic lion algorithm to adaptively form optimal clusters to reduce network energy consumption. To this end, the distance between nodes, the ratio of energy to each load and the distance between nodes and BS are considered to define the similarity, preference, and update responsibility in AP cluster and availability, so as to obtain the optimal number of clusters and initial CHs in each round. In addition, the lion swarm algorithm with a new fitness function considering residual energy and distance to the BS is adopted to optimize the initial CHs, and introduces a chaotic map to improve the convergence of the lion swarm algorithm. The simulation results show that APCLO is superior to APSA and KH in prolonging network lifespan, reducing network energy consumption, and expanding network throughput. Regarding network lifespan performance, APCLO increases 13.5%, 20.1% and 11.2% compared with HSA, KH and APSA protocol, respectively. In reducing energy consumption, APLCO improves by 24.1%, 11.4% and 12.2%, respectively, compared with HSA, KH and APSA. In terms of network throughput, APCLO enhances 5.9%, 15.3%, and 4.7% respectively, compared with APSA, KH and HSA protocols. Moreover, the convergence speed is improved to a certain extent.

REFERENCES

[1] H. Landaluce, L. Arjona, A. Perallos, F. Falcone, I. Angulo, and F. Muralter, "A review of IoT sensing applications and challenges using RFID and wireless sensor networks," *Sensors*, vol. 20, no. 9, pp. 1–18, May 2020.

[2] R. Piyush and C. Siddhartha, "A survey on clustering protocols in wireless sensor network: Taxonomy, comparison, and future scope," *J. Ambient Intell. Humanized Comput.*, vol. 40, pp. 1–47, Jul. 2021.

[3] M. Noureddine, H. A. Zakaria, and E. B. E. A. Abdelbaki, "ECRP: An energy-aware cluster-based routing protocol for wireless sensor networks," *Wireless Netw.*, vol. 26, no. 4, pp. 2915–2928, May 2020.

[4] A. Ghosh and N. Chakraborty, "A novel residual energy-based distributed clustering and routing approach for performance study of wireless sensor network," *Int. J. Commun. Syst.*, vol. 32, no. 7, pp. 1–26, May 2019.

[5] A. Rodriguez, C. Delvallesoto, and R. Velazquez, "Energy-efficient clustering routing protocol for wireless sensor networks based on yellow saddle goatfish algorithm," *Mathematics*, vol. 8, no. 9, pp. 1–17, Sep. 2020.

[6] R. E. Mohamed, A. I. Saleh, M. Abdelrazzak, and A. S. Samra, "Survey on wireless sensor network applications and energy efficient routing protocols," *Wireless Pers. Commun.*, vol. 101, no. 2, pp. 1019–1055, Jul. 2018.

[7] D. W. Sambo, B. O. Yenke, A. Forster, and P. Dayang, "Optimized clustering algorithms for large wireless sensor networks: A review," *Sensor*, vol. 19, no. 2, pp. 1–27, Jan. 2019.

[8] S. Chowdhury and C. Giri, "Energy and network balanced distributed clustering in wireless sensor network," *Wireless Pers. Commun.*, vol. 105, no. 3, pp. 1083–1109, Apr. 2019.

[9] S. K. Singh, P. Kumar, and J. P. Singh, "A survey on successors of LEACH protocol," *IEEE Access*, vol. 5, pp. 4298–4328, 2017.

[10] M. Al-Shalabi, M. Anbar, T.-C. Wan, and Z. Alqattan, "Energy efficient multi-hop path in wireless sensor networks using an enhanced genetic algorithm," *Inf. Sci.*, vol. 500, pp. 259–273, Oct. 2019.

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

[12] A. Al-Baz and A. El-Sayed, "A new algorithm for cluster head selection in LEACH protocol for wireless sensor networks," *Int. J. Commun. Syst.*, vol. 31, no. 1, pp. 1074–5351, Jan. 2018.

[13] A. Wang, D. Yang, and D. Sun, "A clustering algorithm based on energy information and cluster heads expectation for wireless sensor networks," *Comput. Electr. Eng.*, vol. 38, no. 3, pp. 662–671, Jan. 2016.

[14] S. Dutt, S. Agrawal, and R. Vig, "Cluster-head restricted energy efficient protocol (CREEP) for routing in heterogeneous wireless sensor networks," *Wireless Pers. Commun.*, vol. 100, no. 4, pp. 1477–1497, Jun. 2018.

[15] V. S. Chandrawanshi, R. K. Tripathi, and R. Pachauri, "An intelligent low power consumption routing protocol to extend the lifetime of wireless sensor networks based on fuzzy C-means++ clustering algorithm," *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6561–6570, May 2020.

[16] S. A. Sert and A. Yazici, "Increasing energy efficiency of rule-based fuzzy clustering algorithms using CLONALG-M for wireless sensor networks," *Appl. Soft Comput.*, vol. 109, pp. 1568–4946, Sep. 2021.

[17] S. A. Sert, A. Alchihabi, and A. Yazici, "A two-tier distributed fuzzy logic based protocol for efficient data aggregation in multihop wireless sensor networks," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3615–3629, Dec. 2018.

[18] A. Mazinani, S. M. Mazinani, and M. Mirzaie, "FMCR-CT: An energy-efficient fuzzy multi cluster-based routing with a constant threshold in wireless sensor network," *Alexandria Eng. J.*, vol. 58, no. 1, pp. 127–141, Mar. 2019.

[19] A. Panchal and R. K. Singh, "EOCGS: Energy efficient optimum number of cluster head and grid head selection in wireless sensor networks," *Telecommun. Syst.*, vol. 78, no. 1, pp. 1–13, Sep. 2021.

[20] I. Sohn, J.-H. Lee, and S. H. Lee, "Low-energy adaptive clustering hierarchy using affinity propagation for wireless sensor networks," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 558–561, Mar. 2016.

[21] J. Wang, Y. Gao, K. Wang, A. K. Sangaiah, and S. J. Lim, "An affinity propagation-based self-adaptive clustering method for wireless sensor networks," *Sensors*, vol. 19, no. 11, pp. 1–15, Jun. 2019.

[22] H. Cui, L. Wu, Z. He, S. Hu, K. Ma, L. Yin, and L. Tao, "Exploring multidimensional spatiotemporal point patterns based on an improved affinity propagation algorithm," *Int. J. Environ. Res. Public Health*, vol. 16, no. 11, p. 1988, Jun. 2019.

[23] R. Bhardwaj and D. Kumar, "MOFPL: Multi-objective fractional particle lion algorithm for the energy aware routing in the WSN," *Pervas. Mobile Comput.*, vol. 58, Aug. 2019, Art. no. 101029.

[24] P. T. Karthick and C. Palanisamy, "Optimized cluster head selection using krill herd algorithm for wireless sensor network," *Automatika*, vol. 60, no. 3, pp. 340–348, Jul. 2019.

[25] D. Agrawal, S. Pandey, P. Gupta, and M. K. Goyal, "Optimization of cluster heads through harmony search algorithm in wireless sensor networks," *J. Intell. Fuzzy Syst.*, vol. 39, no. 6, pp. 8587–8597, Dec. 2020.

[26] A. Lipare, D. R. Edla, and R. Dharavath, "Fuzzy rule generation using modified PSO for clustering in wireless sensor networks," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 2, pp. 846–857, Jun. 2021.

[27] P. Subramanian, J. M. Sahayaraj, S. Senthilkumar, and D. S. Alex, "A hybrid grey wolf and crow search optimization algorithm-based optimal cluster head selection scheme for wireless sensor networks," *Wireless Pers. Commun.*, vol. 113, no. 2, pp. 905–925, Jul. 2020.

[28] T. A. Alghamdi, "Parametric analysis on optimized energy-efficient protocol in wireless sensor network," *Soft Comput.*, vol. 25, no. 6, pp. 4409–4421, Mar. 2021.

[29] R. Sharma, V. Vashisht, and U. Singh, "eeTMFO/GA: A secure and energy efficient cluster head selection in wireless sensor networks," *Telecommun. Syst.*, vol. 74, pp. 253–268, Feb. 2020.

[30] A. Vinitha, M. S. S. Rukmini, and Dhirajsunehra, "Secure and energy aware multi-hop routing protocol in WSN using Taylor-based hybrid optimization algorithm," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1857–1868, 2022.
[31] Y. Chang, X. Yuan, B. Li, D. Niyato, and N. Al-Dhahir, "A joint unsupervised learning and genetic algorithm approach for topology control in energy-efficient ultra-dense wireless sensor networks," *IEEE Commun. Let.*, vol. 22, no. 11, pp. 2370–2373, Nov. 2021.
[32] X. Cai, Y. Sun, Z. Cui, W. Zhang, and J. Chen, "Optimal LEACH protocol with improved bat algorithm in wireless sensor networks," *KSII Trans. Internet Inf. Syst.*, vol. 13, no. 5, pp. 2469–2490, 2019.
[33] L. Song, Q. Song, J. Ye, and Y. Chen, "A hierarchical topology control algorithm for WSN, considering node residual energy and lightening cluster head burden based on affinity propagation," *Sensors*, vol. 19, no. 13, p. 2925, Jul. 2019.
[34] S. Varshney, C. Kumar, and A. Swaroop, "Lightning-based lion optimization algorithm for monitoring the pipelines using linear wireless sensor network," *Wireless Pers. Commun.*, vol. 117, no. 3, pp. 2475–2494, Apr. 2021.
[35] W. Zhuo and X. Yu, "A particle swarm optimization algorithm based on dynamic adaptive and chaotic search," in *Proc. IOP Conf. Ser. Mater. Sci. Eng.*, 2019, vol. 612, no. 5, Art. no. 052043.

HU HUANG-SHUI received the Ph.D. degree in computer application technology from Jilin University, China, in 2012. He is currently a Professor with the College of Computer Science and Engineering, Changchun University of Technology, China. His research interests include topology control in wireless sensor networks and multifunction vehicle bus networks.

GUO YU-XIN is currently pursuing the master's degree with the College of Computer Science and Engineering, Changchun University of Technology, China. Her research interest includes wireless sensor networks.

WANG CHU-HANG received the master's degree in computer application from Jilin University, China, in 2005. She is currently an Assistant Professor with the College of Computer Science and Technology, Changchun Normal University, China. Her research interests include wireless sensor networks and realtime embedded systems.

GAO DONG is currently pursuing the master's degree with the College of Computer Science and Engineering, Changchun University of Technology, China. His research interest includes security of wireless sensor networks.