Distributed energy-efficient clustering routing protocol for wireless sensor networks using affinity propagation and fuzzy logic

Chu-hang Wang1 · Huang-shui Hu2 · Zhi-gang Zhang2 · Yu-xin Guo3 · Jin-feng Zhang2

Accepted: 25 April 2022 / Published online: 15 June 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Organizing nodes into clusters and forwarding data to the base station (BS) in clustering routing protocols have been widely utilized to improve the energy efficiency, scalability and stability of wireless sensor networks. Making decisions on how many clusters are formed, which nodes are selected as cluster heads (CHs) and who become the relay nodes significantly impact the network performance. Therefore, a distributed clustering routing protocol combined affinity propagation (AP) with fuzzy logic called DAPFL is proposed in this paper, which considers not only energy efficiency but also energy balance to extend the network lifetime. In DAPFL, AP is firstly used to determine the number of clusters and select the best CHs simultaneously based on residual energy, distance between nodes. Then, the optimal next-hop CHs are chosen by using fuzzy logic system with residual energy, data length and distance to BS as descriptors. Simulations in different scenarios are carried out to verify the effectiveness of DAPFL, and the results show that DAPFL exhibits the promising performance in terms of network energy consumption, standard deviation of residual energy, network throughput and lifetime, compared with the up-to-date distributed clustering routing protocols EEFUC, EEFRP, LEACH-AP and APSA.

Keywords WSNs · Affinity propagation · Fuzzy logic · Energy balance and efficiency · Multi-hop routing

1 Introduction
As an important sensing technology of Internet of Things (IoT), WSNs have been widely used in various fields of national economy such as industry, agriculture, construction and transportation (Landaluce et al. 2020). Generally, a WSN consists of thousands of tiny sensor nodes with limited capabilities such as energy, sensing, processing, storage and communication, and the nodes’ energy cannot be replenished for their hostile application environment. During the network operation, each node consumes energy to collect, process and transmit data, and most of the energy is spent on data communication. Therefore, various schemes are adopted to decrease energy consumption so as to support the network in long-term operation, among which clustering routing protocol is considered as the most effective one who groups the nodes into clusters and forwards data to the BS in multi-hop mode (Fanian and Rafsanjani 2019).

In a clustering routing protocol for WSNs, clusters are used to collect data among nodes, and CHs are selected to manage the clusters, aggregate the collected data and forward the fused data to the BS in different ways, as shown in Fig. 1.

It can be seen from Fig. 1 that energy consumption in clustering routing protocols mainly consists of two parts, one is from the communications between cluster members (CMs) and CHs, the other is from the communications between CHs and the BS. Accordingly, forming appropriate clusters and finding optimal routing paths and maintaining clusters properly are common energy saving
methods in clustering routing protocols, which can significantly impact the performance of the network.

In order to form appropriate clusters, the cluster number needs to be determined at first. Traditionally, a fixed value configured by 10% (Richa et al. 2020) or 5% (Noureddine et al. 2020) of the total number of nodes in the network or obtained by geometric calculation (Dutt et al. 2018) is used as the cluster number, which is usually not the most suitable one for a specific application. Therefore, soft computing-based schemes such as harmony search (Alia 2018), cuckoo search (Ghosh and Chakraborty 2019) and yellow saddle goatfish (Rodriguez et al. 2020) algorithms are adopted to search the optimal cluster number, which can evenly distribute energy consumption among clusters.

After determining the number of clusters, CH selection considered as an NP-hard problem (Sambo et al. 2019) is used to find the best nodes as CHs in the clusters, and it is impossible for conventional probability-based (Rawat and Chauhan 2020) or weight-based (Fang and Junfang 2019) approaches whose CHs are selected by a random preset value or a calculated weight value to solve the problem. So soft computing-based approaches are used to obtain the approximate solutions for CH selection due to their local or global search capabilities such as fuzzy logic (Phoemphon et al. 2020), particle swarm optimization (Mohamed et al. 2020), imperialist competitive algorithm (Dehestani and Jamali 2020), genetic algorithms (Kong et al. 2018), moth flame optimization (Richa et al. 2020) and so on. Once the CHs are selected, advertisement messages are broadcast to announce the CH identities, and the normal nodes become CMs by joining the clusters according to different parameters such as the received signal strength, residual energy of the CHs so as to form uniform and energy-efficient clusters. Furthermore, a TDMA schedule is usually utilized to further reduce the intra-cluster energy consumption (Rawat and Chauhan 2020).

Subsequently, the routing mechanism is used to find the optimal routing paths from the source CMs to the BS based on the formed clusters so as to reduce the inter-cluster energy consumption, which is also a NP-hard problem (Sambo et al. 2019). As seen from Fig. 1a, b, a routing path can be denoted by CM-CH-BS (Alia 2018; Rawat and Chauhan 2020) or CM-CH-CH-….CH-BS (Rodriguez et al. 2020; Mohamed et al. 2020), and the former is easy to cause premature death of CHs far away from the BS due to their long-distance data transmission, while the latter is prone to make the CHs near the BS to die early because of their excess burden of data forwarding, which is called hot spot problem usually solved by unequal clustering illustrated in Fig. 1c (Phoemphon et al. 2020). Moreover, the soft computing-based approaches instead of the weight-based approaches selecting the next-hop CH according to some specific parameters such as residual energy, distance and so on (Khoulalene et al. 2018). Al-sodairi and Ouni (2018) are used to find the optimal routing paths so as to reduce and balance the inter-cluster energy consumption, which include fuzzy logic (Jain and Goel 2020), gray wolf optimization (Mohamed et al. 2020), particle swarm optimization (Anand and Pandey 2020), whale optimization (Sakthidasan et al. 2019), genetic algorithm (Bhola et al. 2020) and so on.

Maintaining clusters in round is used to distribute energy consumption among all the nodes, and the fixed round length defined by the time from the beginning of clustering to the end of all source nodes sending data to the BS is the most widely used till now for its simplicity and reliability (Landaluce et al. 2020; Sambo et al. 2019), although variable round length (Fang and Junfang 2019) or adaptive round length (Osamy and Khedr 2020) which can significantly decrease the number of CHs rotation has been validated to be more effective than the fixed round length (Ghosal et al. 2020). Moreover, it is difficult to determine the best round length due to the influence of network dynamics and uncertainties.

As mentioned above, the soft computing-based clustering routing protocols have become the up-to-date schemes to improve energy efficiency and extend network lifetime because of their capabilities of scalability, adaptability and global searching. Especially, fuzzy logic can obtain the best possible solution for CH selection and routing paths finding in WSNs characterized by dynamic and uncertainty (Rajput and Kumaravelu 2020; Balaji et al. 2019).
Moreover, its low complexity is more suitable for applications of WSNs than that of particle swarm optimization, neural network (Taeyoung et al. 2021) and other soft computing-based approaches (Phoemphon et al. 2020). However, a fixed cluster number is usually adopted to form clusters in fuzzy logic-based clustering approaches and a next single-hop CH is found to forward data to the BS in fuzzy logic-based routing approaches, which is almost impossible to form the optimal topology with clusters and to obtain the minimum inter-cluster energy consumption. Moreover, affinity propagation (AP) has been validated to be capable of forming uniform clusters without specifying the cluster number in advance (Wang et al. 2019), and a next CH finding based on the transmitted data length according to hop count also has been verified to further reduce the inter-cluster energy consumption (Nickreay et al. 2015).

Therefore, a distributed clustering routing protocol combined affinity propagation (AP) with fuzzy logic called DAPFL is presented to form clusters and find routing paths in this paper. In DAPFL, without the need to determine the cluster number in advance, AP and fuzzy logic are used to form energy-efficient and balanced clusters and find energy-efficient and balanced routing paths, respectively, so as to maximize the network lifetime. Moreover, the hot spot problem is alleviated by the inter-cluster communication based on appropriate hops, which is depicted in Fig. 1d. The main contributions of this work are summarized as follows.

- AP with novel preference is used to form energy-efficient and balanced clusters, which makes nodes with more residual energy, large average similarity of neighbors have greater chance to be selected as CHs.
- Fuzzy logic system with descriptors residual energy, data length and distance to BS is applied to find the optimal next-hop CHs so as to obtain the energy-efficient and balanced routing paths.
- Performance evaluation is provided to verify the effectiveness of DAPFL compared with the up-to-date algorithms in terms of energy efficiency and balancing, network throughput and lifetime.

The remainder of this paper is organized as follows. The related works are discussed in Sect. 2, and the system model is described in Sect. 3. In Sect. 4, the proposed DAPFL is introduced in detail. In Sect. 5, simulations are performed and results are analyzed in sequence. Finally, Conclusion and future work are provided in Sect. 6.

2 Related works

Clustering routing protocols have been widely used to extend the network lifetime by reducing the network energy consumption since LEACH (low-energy adaptive clustering hierarchy) was proposed in 2000 (Heinzelman et al. 2000). Moreover, numerous variants of LEACH have been presented to overcome its shortcomings (Fanian et al. 2020; Kaur and Mahajan 2018). Comparatively, fuzzy logic-based approaches have been validated to significantly improve the network performance (Mohamed et al. 2020; Rajput and Kumaravelu 2020). Comparatively, fuzzy logic system (FLS) can be used almost in all aspects of clustering routing protocols from competition radius determination, CH selection, cluster formation and routing paths finding due to its adaptive capabilities of uncertainty and dynamic (Phoemphon et al. 2020; Lata et al. 2020). Moreover, distributed methods have been verified to be more suitable for WSNs than centralized ones because of their optimal decision based on only local information (Mazinani et al. 2019). So the state-of-the-art approaches focused on distributed fuzzy logic system are briefly summarized first, which are most relevant in our context.

Fuzzy logic system models the clustering routing process by considering different parameters as descriptors to obtain the best decisions, which mainly consists of fuzzifier, inference engine, knowledge base and defuzzifier (Phoemphon et al. 2020; Rajput and Kumaravelu 2020). Furthermore, all the parameters in FLS such as energy, distance and density have a certain impact on the energy consumption and network lifetime, so the target of FLS is to well integrate various parameters by assigning appropriate membership functions and reason the result by setting different fuzzy rules. In FMCR-CT (Mazinani et al. 2019), two FLS are proposed to elect CHs, in which FLS1 uses parameters residual energy and density of the node as descriptors, and each parameter has 3 linguistic variables whose membership functions are in trapezoidal and triangular form. Nodes with more remaining energy and high density have bigger chance to be elected CHs after inference based on 9 fuzzy rules. Moreover, FLS2 is triggered to elect the CHs once the residual energy of any one CHs elected by FLS1 is less than the threshold value, which has two inputs: remaining energy and distance to CH. Also there are 3 linguistic variables for each parameter and 9 fuzzy rules for inference. As a result, the nodes with more residual energy and less distance to CH are more likely to be elected as CHs. FMCR-CT forms energy-efficient clusters by using FLS, although the hot spot problem is not considered. In UCF (Neamatollahi and Naghijibzadeh 2018), the cluster radius is adjusted to solve the hot spot
Accordingly, in DFCR (Mazumdar and Om 2018), even CHs distribution leads to unbalanced energy consumption. Moreover, its uneven based on residual energy needs large amount of control message and increases collisions. However, random CHs selection only based on residual energy needs large amount of control message and increases collisions. Moreover, its uneven CHs distribution leads to unbalanced energy consumption. Accordingly, in DFCR (Mazumdar and Om 2018), even distributed CHs are elected by FLS with two descriptors residual energy and distance to BS firstly. The member functions of each input parameter have 3 linguistic variables, while that of the output parameter fitness1 has five. Similarly, the cluster radius is obtained by FLS with two descriptors fitness1 and fitness2 with 5 linguistic variables, respectively. Fitness2 is the output of another FLS with descriptors neighbor density and neighbor cost. Neighbor cost denoted by the summation of distance to neighbors. The trapezoidal, triangular and Gaussian membership functions are tested and appropriate ones are picked for the linguistic variables. According to the fuzzy rules (9 for fitness1 and fitness2, 25 for competition radius), nodes closer to the BS with higher residual energy has bigger probability to elected as CHs and CHs closer to the BS with higher residual energy, less neighbors and neighbor cost has bigger cluster radius.

Fuzzy logic system is also utilized to find the optimal forwarder for data transmission. In MLSEEP (Robinson et al. 2019), each CH adopts a FLS to select the best next-hop CH from its neighbor CHs so as to extend the network lifetime and decrease the network overload. Queue length, distance to the BS and residual energy are the three input parameters for FLS, and Queue length denotes the amount of data transferred by a node. There are 3 linguistic variables for each parameters, whose membership functions are in trapezoidal, triangular and Gaussian form. And 27 fuzzy rules are used to make the CHs closer to the BS with higher residual energy and queue length has bigger probability to be selected for routing. Inappropriate CHs are elected to form clusters in MLSEEP by probability based CH selection mechanism. Therefore, unbalanced energy consumption occurs in clusters and the network lifetime is decreased accordingly. In EEFPR (Jain and Goel 2020), a FLS is first proposed to elect CHs with two descriptors residual energy and cost based on the formed even clusters by fuzzy c

Means. Cost is denoted by the summation of distance to neighbors. 3 and 5 linguistic variables are transformed from the two descriptors, respectively, with trapezoidal and triangular membership functions, and 15 fuzzy rules are used for the inference process. Then, the nodes in clusters with more residual energy and small summation of distance to neighbors have bigger probability to be elected as CHs. In addition, the other FLS is used to find the optimal gateways which are nodes with highest residual energy except for CHs in the clusters for data forwarding so as to reduce the burden on CHs. The gateway in a cluster receives the fused data from the CH and then forward the data to its best neighbor gateway determined by the FLS with two descriptors distance to BS and residual energy. Three linguistic variables are transformed from each parameters, all of which use triangular membership functions, and 9 fuzzy rules are set to infer the probability to be selected as next-hop gateways. Fuzzy logic-based clustering and routing in EEFPR forms even clusters and optimal paths in the end, which results in large energy saving to extend the network lifetime. However, it neglects the hot spot problem. Moreover, the CMs join in a cluster only considering received signal strength from their neighbor CHs, which leads to reduced network lifetime due to the unbalanced energy consumption in clusters. Consequently, in EEFUC (Phoemphon et al. 2020), 4 fuzzy logic systems with different descriptors are presented to not only determine competition radius, elect CHs, select the next-hop CH but also join an appropriate cluster for each CM. Residual energy, node density and distance to BS are used as fuzzy input parameters to determine the competition radius. And the CHs are determined by their residual energy of nodes, and a FLS is invoked to select the best one as the final CHs while their residual energy is the same, whose input parameters are node density and distance to BS. Once the CHs are selected, another FLS with input parameters residual energy and distance to CH is utilized to make a CM join a specific CH, then the optimal clusters are formed. Finally, a FLS is called by each CH to find the most suitable next-hop CH, which makes decision based on the parameters residual energy of neighbor CH, distance to neighbor CH and DOP which is the distance to the line from CH to BS. All the parameters are transformed to 3 linguistic variables, and tests are performed to select the superior membership functions including trapezoidal, triangular, Gaussian and hybrid for different linguistic variables. In addition, the fuzzy rules for each FLS are the number of parameters power 3. In this way, EEFUC can extend the network lifetime largely in various scenarios. The reviewed approaches are summarized in Table 1.
| Protocol (year) | Cluster size | Hot spot problem | FLS | Purpose | Parameters | Linguistic variables | Number of rules | Target |
|----------------|--------------|------------------|-----|---------|------------|---------------------|-----------------|--------|
| FMCR-CT (2018) | Unequal Not considered | CH Selection | Input | FLS1 Residual energy | Density of the node, FLS2 Remaining energy, Distance to CH | Low, average, high, Few, average, high, Low, average, high, Very little, little, average, far, too far | 9 | Energy efficiency |
| Output | FLS1 Chance, FLS2 Chance | Low, average, high, Little, average, high | 9 |
| UCF (2018) | Unequal Considered | Cluster radius calculation | Input | Distance to BS | Local density | Low, medium, high, Very high, high, rather high, medium high, medium, medium low, rather low, low, very low | | Energy efficiency |
| Output | Cluster radius | Very high, high, rather high, medium high, medium, medium low, rather low, low, very low | 9 |
| DFCR (2018) | Unequal Considered | CH Selection | Input | Residual energy, Distance to BS | Fitness1 | Small, average, high, Near, moderate, distant, Medium, weak, very weak, strong, very strong | 9 | Energy efficiency, Energy balance |
| Output | Fitness2, FLS3 Cluster radius | Medium, weak, very weak, strong, very strong, Very high, high, low, medium, very low, Long, Med long, medium, Med small, small, V small, V V small, fairly long, V long | |
| MLSEEPEP (2019) | Equal | Not considered | Next-hop CH finding | Input | Queue length, distance to the BS, residual energy | Small, normal, large | 27 | Energy efficiency |
| Output | Probability | Small, normal, large | 27 |
| EEFRP (2020) | Equal | Not considered | CH Selection | Input | Residual energy | Very low, low, medium, high, very high, High, low, medium, Very low, low, high, medium, very high, Low, medium, high | 15 | Energy efficiency |
| | | | | Output | Possibility | Medium, high, very high, low, very low | | |
| | | | | | | | | |
The optimal number of clusters is not considered in all the approaches reviewed above. AP is a new algorithm for clustering without specifying the number of clusters in advance. It has been used to solve the problem with large-scale data only through setting a few parameters, which is superior to the geometric and soft computing-based schemes (Cui et al. 2019; Liu et al. 2019). In AP algorithms, the similarity matrix is used to represent the correlations between nodes followed by iterative simple message exchanges to obtain the optimal clusters, which has the most important impact on the performance of clusters. In LEACH-AP (Sohnet al. 2016), the negative energy consumption of the link between nodes is used to define the similarity between two nodes, so as to make the nodes with large value belong to the same cluster. Moreover, preference denoted by the self-similarity is defined to make nodes with large value become CHs. Although the optimal clusters can be formed and the optimal number of clusters is not required in LEACH-AP, the network energy efficiency will be decreased largely due to the long distance transmission in clusters. Therefore, in APSA (Wang et al. 2019), the negative Euclidean distance is used to represent the similarity between two nodes, so as to make nodes close to each other be a cluster. The clusters formed in APSA are more uniform than those in LEACH-AP; however, the fixed preference makes it difficult to select the best CHs.

### 3 System model

In DAPFL, $n$ nodes with limited energy are randomly scattered in the network with a target region $M \times Mm^2$ as depicted in Fig. 1d, and clusters like in LEACH (Heinzelman et al. 2000) are used to organize the nodes with unique IDs. In a cluster, a specific node is selected as CH which is responsible for management of the cluster including forming the cluster, receiving data from its CMs, aggregating and sending data directly or indirectly to the BS, maintaining the cluster and so on. At the same time, CMs directly communicate with their CHs only. Moreover, the following assumptions are considered in the presented network model.

- All the nodes are static including the BS.
- All the nodes are homogeneous with the same capabilities of sensing, processing, storage, communication and initial energy except the BS.
- Symmetric links are used for communication between any two nodes.
- The distance between two nodes can be obtained according to the received signal strength.

In order to calculate the energy consumption of each node, the first-order radio model like in (Wang et al. 2019; Heinzelman et al. 2000) is used in this paper. When a node

### Table 1 (continued)

| Protocol (year) | Cluster size | Hot spot problem | FLS Purpose | Parameters | Linguistic variables | Number of rules | Target |
|----------------|--------------|------------------|-------------|------------|----------------------|----------------|--------|
| EEFUC (2020)   | Unequal     | Considered       | CH Selection | Input: Distance to BS, Node density | Near, medium, far, Low, medium, high | 9 | Energy balance |
|                |              |                  |             | Output: CH Chance | VVH, VH, H, M, HM, RM, L, VL, VVL |
|                |              |                  | Cluster radius calculation | Input: Distance to BS, Remaining energy, node density | Near, medium, far, Low, medium, high | 27 |
|                |              |                  |             | Output: Competition radius | 12-2XS, XS,S,M,G,XG,2-12XG |
|                |              |                  | CM joining | Input: Distance to CH, Remaining energy | Far, medium, near, Low, medium, high | 9 |
|                |              |                  |             | Output: CM Chance | VVL, VL, L, RM, M, HM, H, VH, VVH |
|                |              |                  | Next-hop CH finding | Input: Remaining energy, Delay distance, distance to optimal point | Low, medium, high, Far, medium, near | 27 |
|                |              |                  |             | Output: nextCH Chance | 12-2VL, VL, L, M, H,VH, 2-12VH |
where $E_{elec}$ denotes the energy consumption for transmitting or receiving 1-bit data, $\epsilon_f$ and $\epsilon_mp$ are the amplifier coefficients of free space and multi-path fading, respectively, $d_0$ is the threshold distance given by $d_0 = \sqrt{\epsilon_f / \epsilon_mp}$.

At the same time, the amount of energy consumption of receiving $l$ - bits data for node $i$ from node $j$ is given by

$$E_{Rij} = l \ast E_{elec}.$$  \hspace{1cm} (2)

And the amount of energy consumption for aggregating $l$ - bits data is given by

$$E_{DA} = l \ast E_{pDb} \hspace{1cm} (3)$$

where $E_{pDb}$ is the energy consumption for 1-bit data fusion.

## 4 Proposed protocol

In DAPFL, affinity propagation is used for clustering, and makes the nodes with more residual energy and closer to cluster center be CHs. Based on the determined CHs, fuzzy logic system with descriptors residual energy, data length and distance to BS is adopted to find the optimal routing paths for CHs, which is discussed in detail as below.

### 4.1 Affinity propagation based clustering

To minimizing the energy consumption of a cluster, it is necessary to minimize the distance between any CM and the CH; therefore, the absolute value of the difference of distance to the BS of node $i$ and $j$ is used to calculate their similarity, which is expressed as follows:

$$s(i, k) = -|d(i, BS) - d(k, BS)|, \ i, k \in [1, n], i \in N_k, i \neq k \hspace{1cm} (4)$$

where $d_{i_oBS}$ denotes the distance to the BS of node $i$. Moreover, $s(i,k)$ sets to negative infinity when node $i$ cannot directly communicate with node $k$. In addition, the preference $s(k,k)$ indicates that node $k$ will be selected as CH is given by

$$s(k, k) = \frac{1}{2 - \alpha} + \beta \hspace{1cm} (5)$$

where:

- $\alpha$ denotes the normalized average similarity of its neighbors for node $k$, which is shown as:

$$\alpha = \frac{\text{Avg}S_k - \text{Avg}S_{\min}}{\text{Avg}S_{\max} - \text{Avg}S_{\min}} \hspace{1cm} (6)$$

$	ext{Avg}S_k = \sum_{i \in N_k}s(i,k)/N_k$ represents the average similarity of its neighbors for node $k$, $|N_k|$ is the number of neighbors of node $k$, and $\text{Avg}S_{\min}$ and $\text{Avg}S_{\max}$ are the minimum and maximum of $\text{Avg}S_i, i \in N_k$.

- $\beta$ denotes the normalized average residual energy for node $k$, which can be described as:

$$\beta = \frac{\text{Eravg}_k - \text{Eravg}_{\min}}{\text{Eravg}_{\max} - \text{Eravg}_{\min}} \hspace{1cm} (7)$$

$\text{Eravg}_k = \sum_{i \in N_k}\frac{\text{Eres}_k}{|N_k|}$, $\text{Eres}_k$ is the residual energy of node $k$. $\text{Eravg}_{\min}$ and $\text{Eravg}_{\max}$ are the minimum and maximum of $\text{Eravg}_i = \sum_{i \in N_k}\frac{\text{Eres}_i}{|N_k|}, i \in N_k$. As seen from Eq. (5), the nodes with more residual energy and large average similarity of neighbors have greater chance to be selected as CHs. Moreover, residual energy as well as availability $a(i,k)$ is used to update responsibility $r(i,k)$ using Eqs. (8) and (9), respectively. The former means the degree to node $k$ selected as the CH of node $i$, and the latter reflects the appropriate degree of node $i$ to select $k$ as its CH.

$$r(i,k) = \frac{\text{Eres}_k}{d_{ik}}s(i,k) - \max_{k' \neq k} (a(i,k') + s(i,k')) \hspace{1cm} (8)$$

$$a(i,k) = \begin{cases} \sum_{i \neq k} \max_{i \neq k} \left\{ \frac{\text{Eres}_k}{d_{ik}} \ast r(i', k) \right\}, & \text{if } i = k \\ \min_{i \neq k} \sum_{i \neq k} \max_{i \neq k} \left\{ \frac{\text{Eres}_k}{d_{ik}} \ast r(i', k) \right\}, & \text{if } i \neq k \end{cases} \hspace{1cm} (9)$$

Nodes with more energy and smaller average distance to neighbors are more likely to be selected as CHs. Moreover, the initial value of $a(i,k)$ is set to zero. By using the corresponding values in the last iteration, the updating continues until the iteration number reaches the preset threshold or the estimated preferences stay the same for a certain number of iterations (Wang et al. 2019; Liu et al. 2019). And then the nodes meeting with $r(k,k) + a(k,k) > 0, k \in [1, n]$ are selected as CHs. The other nodes determine their relevant CHs according to the similarity values. Moreover, TDMA mechanism like in Heinzelman et al. (2000) is used to save energy consumption for intra-cluster communication.
4.2 Fuzzy logic-based routing

Once the clusters are formed by running AP, FLS is utilized to accomplish optimal routing in DAPFL, which is a completely distributed scheme to make decisions only based on the local information for each CH. Moreover, different uncertainties are handled by FLS with various parameters in DAPFL like the traditional fuzzy routing methods (Phoemphon et al. 2020; Jain and Goel 2020) so as to make it applied in practical WSNs. As is known to all, many parameters have significant impact on the performance of routing in WSNs, and they should be considered carefully for the optimal decisions. Therefore, in DAPFL, residual energy (RE), distance to BS (DB) and data length (DL) are considered as input parameters for a Mamdani fuzzy logic system like in Phoemphon et al. (2020) and Jain and Goel (2020) whose block diagram is shown in Fig. 2.

Different from the traditional approaches whose clusters nearer to the BS have smaller cluster radius to deal with the hot spot problem, DAPFL balances the energy consumption of CHs by determining appropriate parameters for FLS. The specifics and significance of the input parameters are described as follows:

- **RE** Residual energy of the candidate next-hop CH means the current energy level, so high residual energy indicates that the candidate next-hop CH has enough energy for data forwarding. In other words, the candidate next-hop CH with more RE has higher chance to be select as the next-hop CH.

| Parameters | Linguistic variables | Membership functions |
|------------|----------------------|----------------------|
| RE         | L(Low), H(High)      | Trapezoidal          |
|            | M(Medium)            | Triangular           |
| DB         | N(Near), F(Far)      | Trapezoidal          |
|            | M(Medium)            | Triangular           |
| DL         | S(Short), L(Long)    | Trapezoidal          |
|            | M(Medium)            | Triangular           |
| Chance     | L(Low), M(Medium), H(High), V(L(Very Low), VH(Very High)) | Triangular |
|            | VV(L(Very Very Low), VVH(Very Very High)) | Trapezoidal |

Fig. 2 Block diagram of the fuzzy logic system

Fig. 3 Schematic diagram of parameter combination

Table 2 Input and output parameters, linguistic variables and their membership functions
Distance to BS of the candidate next-hop CH is used to measure how far the candidate is from the BS, as shown in Fig. 3. Accordingly, lower distance to BS means the candidate next-hop CH consumes less energy to communicate with the BS.

Data length of the source CH represents the amount of data to be forwarded by the candidate next-hop CH, which is used to adjust the hops between the source CH and the candidate next-hop CH, as shown in Fig. 3. The source CH with lower DL has higher chance to select the candidate next-hop CH with lower DB as the next-hop CH.

Seen from Fig. 3, only the CHs with smaller DB than the source CH1 are considered to be candidate next-hop CHs so as to find the optimal next-hop CH quickly. According to the distance to the source CH1, the hop for each candidates is determined. One with the smallest distance is the first hop (CH2), the next is the second hop (CH3, CH4) and so on. The target hop of the next-hop CH is mainly determined by RE and DL. Based on the target hop, the candidate next-hop CH with lower DB has higher chance to be selected as the next-hop CH.

Before the FLS makes final decision which candidate next-hop CH with appropriate RE, DB and DL becomes the optimal next-hop CH, its input and output parameters in crisp data are converted to suitable linguistic variables whose membership functions are derived accordingly. The detail linguistic variables and parameters are tabulated in Table 2. Moreover, all the values of the linguistic variables are normalized in the range from 0 to 1 by using the min–

**DB** Distance to BS of the candidate next-hop CH is used to measure how far the candidate is from the BS, as shown in Fig. 3. Accordingly, lower distance to BS means the candidate next-hop CH consumes less energy to communicate with the BS.

**DL** Data length of the source CH represents the amount of data to be forwarded by the candidate next-hop CH, which is used to adjust the hops between the source CH and the candidate next-hop CH, as shown in Fig. 3. The source CH with lower DL has higher chance to select the candidate next-hop CH with lower DB as the next-hop CH.

Seen from Fig. 3, only the CHs with smaller DB than the source CH1 are considered to be candidate next-hop CHs so as to find the optimal next-hop CH quickly. According to the distance to the source CH1, the hop for each candidates is determined. One with the smallest distance is the first hop (CH2), the next is the second hop (CH3, CH4) and so on. The target hop of the next-hop CH is mainly determined by RE and DL. Based on the target hop, the candidate next-hop CH with lower DB has higher chance to be selected as the next-hop CH.

Before the FLS makes final decision which candidate next-hop CH with appropriate RE, DB and DL becomes the optimal next-hop CH, its input and output parameters in crisp data are converted to suitable linguistic variables whose membership functions are derived accordingly. The detail linguistic variables and parameters are tabulated in Table 2. Moreover, all the values of the linguistic variables are normalized in the range from 0 to 1 by using the min–
max normalization method. Then, the linguistic variables are processed by the inference engine to construct a functional mapping between the input and output variables based on the knowledge base consisting of if–then rules. Due to their significant affection on the inference results, the membership functions for different variables and rules are explored by a large number of tests using the trial-and-error method. The final chosen membership and fuzzy rules are shown in Fig. 4 and Table 3, respectively.

The widely used center of area (COA) method is used for defuzzification in order to convert the fuzzy output Chance to crisp output. After the candidate next-hop CHs calculate the Chance of becoming the optimal next-hop CH, each of them sends the Chance value to the source CH, and the source CH selects the one with the highest Chance value as its next-hop CH.

### 5 Performance evaluation

The performance of DAPFL is evaluated in this section based on MATLAB R2018a, compared with the up-to-date clustering routing protocols EEFUC (Phoemphonet al. 2020), EEFRP (Jain and Goel 2020), LEACH-AP (Sohnet al. 2016) and APSA (Wang et al. 2019). The static nodes are randomly scattered over the sensing area of the network, and two scenarios are constructed to provide the test environment. The scenario #1 has a small dimension of 200 × 200 m^2 while the scenario #2 has a higher dimension which covers 500 × 500 m^2. In addition, 100 and 1000 nodes are deployed in each scenario. Moreover, the BS locates at the position away from the sensing area for practical applications. The initial energy of all the nodes is the same. The detail parameter settings are given in Table 4.

The total energy consumption in each round is usually used to judge the network energy efficiency. The less energy consumption is, the higher the network energy efficiency. Thus, the total energy consumption per round is tested in the different scenarios, and the results are shown in Fig. 5.

Seen from Fig. 5a, b, it is obvious that the total energy consumption increases with the rounds going in all cases, and DAPFL increases more slowly than the other four protocols. When 90% of the energy is consumed, the rounds for DAPFL occur at 1207 and 552, while 1113 and 488 for EEFUC, 1051 and 443 for EEFRP, 892 and 391 for LEACH-AP, 988 and 428 for APSA. The average energy consumption of DAPFL has decreased by 8.45%, 13.11% over EEFUC, 14.84%, 24.6% over EEFRP, 35.62%, 41.18% over LEACH-AP, and 22.17%, 28.97% over APSA, respectively.

However, the total energy consumption might be consumed by only a small number of nodes, which leads to unbalanced energy consumption. So the standard deviation of residual energy of all the nodes are tested, and the less

---

**Table 3** Fuzzy if–then rules

| Rule no | Input | Output |
|---------|-------|--------|
| RE | DB | DL | Chance |
| 1 | L | N | S | M |
| 2 | L | N | M | L |
| 3 | L | L | M | VVL |
| 4 | L | M | S | L |
| 5 | L | M | M | VVL |
| 6 | L | M | S | VVL |
| 7 | L | F | S | VL |
| 8 | L | F | M | VVL |
| 9 | L | F | L | VVL |
| 10 | M | N | S | VH |
| 11 | M | N | M | M |
| 12 | M | N | L | M |
| 13 | M | M | L | M |
| 14 | M | M | M | L |
| 15 | M | F | S | H |
| 16 | M | M | M | M |
| 17 | M | M | L | M |
| 18 | H | M | S | VH |
| 19 | H | N | S | VVH |
| 20 | H | N | M | VH |
| 21 | H | N | L | H |
| 22 | H | M | M | VH |
| 23 | H | F | M | M |
| 24 | H | M | L | H |
| 25 | H | F | S | VH |
| 26 | H | F | M | H |
| 27 | H | F | L | M |

**Table 4** Simulation parameter settings

| Number of nodes | 100,1000 | 100,1000 |
|-----------------|----------|----------|
| Initial energy  | 1 J      | 1 J      |
| $E_{elec}$     | 50 (nJ/bit) | 50 (nJ/bit) |
| $e_{fs}$       | 10 (pJ/bit/m^2) | 10 (pJ/bit/m^2) |
| $e_{mp}$       | 0.0013 ((pJ/bit/m^4)) | 0.0013 ((pJ/bit/m^4)) |
| $d_0$          | 87 m     | 87 m     |
| Data packet size | 4000 bits | 4000bits |
| Control packet size | 200 bits | 200bits |
| Area           | 200 m*200 m | 500 m*500 m |
| BS location    | x = 100, y = 100 | x = 600, y = 600 |
the standard deviation of residual energy, the better the network balance. The results are depicted in Fig. 6.

It can be seen from Fig. 6, whether in Scenario #1 or #2, the standard deviation of residual energy for DAPFL is the smallest, which indicates it can exhibit the best performance in network energy balance. The standard deviation of residual energy for DAPFL is 12.5%, 19.54%, 30.69%, 30.56% lower than those of the other four algorithms in Scenario #1 with 100 node, while 31.47%, 44.33%, 49.77%, 65.17% in Scenario #1 with 1000 node. Similarly, the standard deviation of residual energy for DAPFL has decreased by 13.06% and 10% over EEFUC, 30.1%, 21.89% over EEFRP, 47.81%, 43.32% over LEACH-AP, and 36.8%, 36.79% over APSA in Scenario #2 with 100 and 1000 nodes, respectively.

Despite all this, more energy might be wasted for invalid data transmission, thus only the useful data transmission from CHs to the BS is considered for the network throughput which is used to measure the quality of service (QoS) of the network. Furthermore, it directly reflects the CHs distribution and energy balance. The larger the network throughput is, the better the QoS of the protocols. The comparison results are illustrated in Fig. 7.

As can be seen from Fig. 7, for all the protocols, their network throughput increases with the number of rounds in all scenarios. Moreover, DAPFL has the highest network throughput, compared with EEFUC, EEFRP, APSA and LEACH-AP. It outperforms EEFUC by 9.38%, 20.43%, 27.45%, 11.92%, EEFRP by 14.51%, 33.94%, 45.91%, 28.54%, LEACH-AP by 35.74%, 62.83%, 76.2%, 68.83%.

Fig. 5 Comparison of the total energy consumption versus rounds

(a) Energy consumption in Scenario #1 with 100 nodes
(b) Energy consumption in Scenario #1 with 1000 nodes
(c) Energy consumption in Scenario #2 with 100 nodes
(d) Energy consumption in Scenario #2 with 1000 nodes
and APSA by 27.99%, 42.22%, 80.98% and 28.94% in the four cases of Scenario #1 and #2, respectively.

Extending the network lifetime as much as possible is the main goal for all clustering routing protocols, and the network lifetime has been defined in different ways. Here, the widely used definition of the network lifetime is adopted to verify the performance of the protocols, namely the number of rounds when all the nodes die. Moreover, the first node die (FND), half node die (HND) and last node die (LND) are also considered to analyze the network lifetime. The simulation results are shown in Fig. 8 and Table 5.

Figure 8 shows that DAPFL has the longest network lifetime in all the simulation scenarios. The average network lifetime DAPFL is enhanced by 4.57%, 71.33%, 111.43%, 74.89% over EEFUC, 5.75%, 114.67%, 203.33%, 172.5% over EEFRP, 13.1%, 161.85%, 380%, 218.57% over LEACH-AP and 9.83%, 130.63%, 298.33%, 218.57% over APSA, respectively. Moreover, DAPFL also exhibits superior performance in FND, HND and LND. It can be seen.

from Table 5 that FND of DAPFL occurs at round 1201, 28, 1 and 1, followed by EEFUC at round 896, 9, 1 and 1, EEFRP at round 960, 14, 1 and 1, LEACH-AP at round 789, 22, 1 and 1, APSA at round 866, 18, 1 and 1. For HND, DAPFL occurs at round 1350, 498, 35 and 237, while EEFUC at round 1226, 449, 31 and 181, EEFRP
at round 1168, 388, 27 and 154, LEACH-AP at round 985, 363, 21 and 126, APSA at round 1061, 372, 23 and 156. Similarly, DAPFL has the longest LND with values of 1466, 1004, 86 and 435, and the other four protocols follow in the order of EEFUC with values 1451, 833, 75 and 418, EEFPR at round 1349, 806, 67 and 368, LEACH-AP at round 1109, 633, 58 and 337, APSA at round 1213, 721, 62 and 352, respectively.

The simulation results of Figs. 5, 6, 7, 8 and Table 5 show that DAPFL achieves superior performance compared with the other protocols EEFUC, EEFPR, LEACH-AP and APSA in different scenarios most of the time, because DAPFL can adaptively form optimal number of equal clusters using AP, and find the best path for each CH using FLS. Different from the FLS in EEFUC and EEFPR, data length is considered to determine the hops so as to find the appropriate next-hop CH. Although LEACH-AP and APSA can also adaptively construct optimized clusters by AP, the single-hop communication between CHs and the BS leads to a large number of long-distance data transmission, which greatly increases the network energy consumption, especially in large-scale network. Moreover, LEACH-AP generates intra-cluster long distance data transmission between CMs and the CH due to its only

Fig. 7 Comparison of the network throughput versus rounds
considering residual energy during the course of clustering. Thus, LEACH-AP exhibits the worst overall performance. In EEFUC and EEFRP, FLS is used to not only select CHs but also find the next-hop CHs, so both of them exhibit better performance than LEACH-AP and APSA. Particularly, FLS is also adopted to calculate the cluster radius and complete the CMs joining in EEFUC so as to form optimal clusters and alleviate the hot spot problem, therefore it outperforms EEFRP. However, hop-by-hop data forwarding is prone to increase the network energy consumption as well as the end-to-end delay, especially in the case of small amount of data forwarding.

6 Conclusion

In this paper, an energy-efficient and load balanced clustering routing protocol DAPFL based on affinity propagation and fuzzy logic is proposed to adaptively form clusters and find the optimal routing paths. To this end, residual energy and distance are considered to define the similarity, preference and update the responsibility, availability in AP clustering so as to make the best nodes located at the center of clusters be CHs. And then fuzzy logic system is used to find the optimal next-hop CH for each CH. The parameters residual energy, distance to BS and data length are carefully determined to calculate the Chance of being the final next-hop CHs. Meanwhile, the hot spot problem is
alleviated by balancing the energy consumption among the CHs based on data length, which is different from the traditional methods by adjusting the size of all clusters based on distance to the BS. Simulations are conducted to prove the effectiveness of DAPFL, and the results show that it is superior to EEFUC, EEFRP, LEACH-AP and APSA. With respect to the network energy consumption, the average values of DAPFL in all the scenarios are 10.045%, 19.2975%, 38.69% and 25.7825% lower than those of EEFUC, EEFRP, LEACH-AP and APSA, respectively. The average standard deviation of residual energy in all the scenarios for DAPFL is 16.76%, 28.94%, 42.9% and 42.33% lower than those of the other four algorithms. At the same time, DAPFL outperforms EEFUC, EEFRP, LEACH-AP and APSA in the average network throughput for all the scenarios by 17.3%, 30.73%, 52.83% and 45.03%. As for the network lifetime, DAPFL enhances it by 8.37% over EEFUC, 19.95% over EEFRP, 42.04% over LEACH-AP and 30.06% over APSA in all the Scenarios. In the future, mobile nodes including the BS will be considered for more extensive practical applications. Besides, blending additional parameters for the FLS such as secure link and packet loss rate will also be explored in depth so as to maximize the network lifetime. Especially, the deep learning methods will also being considered to further improve the overall performance of the network.

**Authors’ contributions** WC was involved in conceptualization and writing—original draft preparation; HH helped in methodology; ZZ contributed to formal analysis and investigation; GY was involved in writing—review and editing; ZJ helped in writing—review and editing.

**Funding** The author would like to thank the National Natural Science Foundation of China for supporting this work by Grant Code: 61803044, the science and technology development project of Jilin province by Grant Code: 20210201051GX and the education department project of Jilin province by Grant Code: JJKH20220686KJ.

### Declarations

**Informed consent** All authors read and approved the final manuscript.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** Not applicable.

### References

Alia OM (2018) A dynamic harmony search-based fuzzy clustering protocol for energy-efficient wireless sensor networks. Ann Telecommun 73(5):353–365

Al-sodairi S, Ouni R (2018) Reliable and energy-efficient multi-hop LEACH-based clustering protocol for wireless sensor networks. Sustain Comput-Inf Syst 20(2018):1–13

Anand V, Pandey S (2020) New approach of GA-PSO-based clustering and routing in wireless sensor networks. Int J Commun Syst 33(16):1–20

Balaji S, Julie EG, Robinson YH (2019) Development of fuzzy based energy efficient cluster routing protocol to increase the lifetime of wireless sensor networks. Mobile Netw Appl 24(2):394–406

Bhola J, Soni S, Cheema GK et al (2020) Genetic algorithm based optimized leach protocol for energy efficient wireless sensor networks. J Ambient Intell Humaniz Comput 11(3):1281–1288

Cui HF, Wu L, He ZJ et al (2019) Exploring multidimensional spatiotemporal point patterns based on an improved affinity propagation algorithm. Int J Environ Res Public Health 16(11):1–19

Dehestani F, Jamali MAJ (2020) Load balanced clustering based on imperialist competitive algorithm in wireless sensor networks. Wireless Pers Commun 112(1):371–385

Dutt S, Agrawal S, Vig R (2018) Cluster-head restricted energy efficient protocol (CREEP) for routing in heterogeneous wireless sensor networks. Wireless Pers Commun 100(4):1477–1497
Fang Z, Junfang W (2019) An energy-efficient unequal clustering routing protocol for wireless sensor networks. Int J Distr Sens Netw 15(9):1–15

Fanian F, Rafsanjani MK (2019) Cluster-based routing protocols in wireless sensor networks: a survey based on methodology. J Netw Comput Appl 142(2019):111–142

Ghosal A, Halder S, Das SK (2020) Distributed on-demand clustering algorithm for lifetime optimization in wireless sensor networks. J Parallel Distrib Comput 141(2020):129–142

Ghosh A, Chakraborty N (2019) A novel residual energy-based distributed clustering and routing approach for performance of wireless sensor network. Int J Commun Syst 32(7):1–26

Heinzelman W R, Chandrakasan A, Balakrishnam H,(2000) Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Annual Hawaii international conference on system sciences. Piscataway, NJ: IEEE: 3 senate005–3014

Jain A, Goel AK (2020) Energy efficient fuzzy routing protocol for wireless sensor networks. Wireless Pers Commun 110(3):1459–1474

Kaur S, Mahajan R (2018) Hybrid meta-heuristic optimization based energy efficient protocol for wireless sensor networks. Egypt Inf J 19(3):145–150

Khoualene N, Bouallouche ML, Aissani D et al (2018) Clustering with load balancing-based routing protocol for wireless sensor networks. Wireless Pers Commun 103(3):2155–2175

Kong LP, Pan JS, Snasel V et al (2018) An energy-aware routing protocol for wireless sensor network based on genetic algorithm. Telecommun Syst 67(3):451–463

Landaluce H, Arjona L, Perallos A et al (2020) A review of IoT sensing applications and challenges using RFID and wireless sensor networks. Sensors 20(9):1–18

Lata S, Mehfuz S, Urooj S et al (2020) Fuzzy clustering algorithm for enhancing reliability and network lifetime of wireless sensor networks. IEEE Access 8(2020):66013–66024

Liu YC, Liu JC, Jin YC et al (2019) An affinity propagation clustering based particle swarm optimizer for dynamic optimization. Knowl-Based Syst 195(2020):1–14

Mazinani A, Mazinani SM, Mirzaie M (2019) FMCR-CT: An energy-efficient fuzzy multi cluster-based routing with a constant threshold in wireless sensor network. Alex Eng J 58:127–141

Mazumdar N, Om H (2018) Distributed fuzzy approach to unequal clustering and routing algorithm for wireless sensor networks. Int J Commun Syst 31(12):1–23

Mohamed E, Rundar R, Mohammad H et al (2020) Swarm intelligence based energy efficient protocol for multipath routing protocol for sustainable wireless sensor networks. Int J Distr Sens Netw 16(9):1–12

Neamatollahi P, Naghibzadeh M (2018) Distributed unequal clustering algorithm in large-scale wireless sensor networks using fuzzy logic. J Supercomput 74(6):2329–2352

Nickreay M, Afzali KA, Janti R (2015) Simultaneous power control and power management algorithm with sector-shaped topology for wireless sensor networks. EURASIP J Wirel Commun Netw 118(2015):1–16

Noureddine M, Zakaria HA, Abdelbaki EBEA (2020) ECRP: an energy-aware cluster-based routing protocol for wireless sensor networks. Wireless Netw 26(4):2915–2928

Osamy W, Khedr AM (2020) Adaptive and dynamic mechanism for round length determination in cluster based wireless sensor networks. Wireless Pers Commun 114(2):1155–1175

Phoemphon S, Soin C, Aintongkham P et al (2020) An energy-efficient fuzzy-based scheme for unequal multihop clustering in wireless sensor networks. J Ambient Intell Humaniz Comput 5(2020):1–23

Rajput A, Kumaravelu VB (2020) FCM clustering and FLS based CH selection to enhance sustainability of wireless sensor networks for environmental monitoring applications. J Ambient Intell Humaniz Comput 6(2020):1–21

Rawat P, Chauhan S (2020) Probability based cluster routing protocol for wireless sensor network. J Ambient Intell Humaniz Comput 10(2020):1–13

Richa S, Vasudha V, Umang S (2020) eeTMFO/GA: a secure and energy efficient cluster head selection in wireless sensor networks. Telecommun Syst 74(2020):253–268

Robinson YH, Julie EG, Kumar R et al (2019) Probability-based cluster head selection and fuzzy multipath routing for prolonging lifetime of wireless sensor networks. Peer-to-Peer Network Appl 12(5):1061–1075

Rodriguez A, Delvallesoto C, Velazquez R (2020) Energy-efficient clustering routing protocol for wireless sensor networks based on yellow saddle goatfish algorithm. Mathematics 8(9):1–17

Sakthidasan K, Vasudevan N, Diderot PKG, et al (2019) WOAPR: an affinity propagation based clustering and optimal path selection for time critical wireless sensor networks. IET Networks 8(2):100–106

Sambo DW, Yenke BO, Foerster A, et al (2019) Optimized clustering algorithms for large wireless sensor networks: a review. Sensors, 19(2):1–27.

Singh SK, Kumar P, Singh JP (2018) A survey on successors of LEACH protocol. IEEE Access 5(2017):4298–4328

Sohn I, Lee JH, Lee SH (2016) Low-Energy adaptive clustering hierarchy using affinity propagation for wireless sensor networks. IEEE Commun Lett 20(3):558–561

Taeyoung K, Luiz FV, Kyujin C et al (2021) Machine learning for advanced wireless sensor networks: a review. IEEE Sens J 21(11):12379–12397

Wang J, Gao Y, Wang K et al (2019) An affinity propagation-based self-adaptive clustering method for wireless sensor networks. Sensors 19(11):1–15

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.