A green cluster-based routing scheme for large scale wireless sensor networks

A Green Cluster-based Routing Scheme for Large Scale Wireless Sensor Networks

Prasenjit Chanak[*,a], Indrajit Banerjee[b], R. Simon Sherratt[c]

[a]*Department of Information and Communication Technology, Atal Bihari Vajpayee Indian Institute of Information Technology & Management, Gwalior-474015, India*
[b]*Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-03, India*
[c]*Department of Biomedical Engineering, University of Reading, Reading, Berkshire, RG6 6AY, UK*

Abstract — In Wireless Sensor Networks (WSNs), clustering has been shown to be an efficient technique to improve scalability and network lifetime. In clustered networks, clustering creates unequal load distribution among Cluster Heads (CHs) and Cluster Member (CM) nodes. As a result, the entire network is subject to premature death because of the deficient active nodes within the network. In this paper, we present clustering-based routing algorithms that can balance out the trade-off between load distribution and network lifetime "green cluster-based routing scheme". This paper proposes a new energy aware green cluster-based routing algorithm to preventing premature death of large scale dense WSNs. To deal with the uncertainty present in network information, a fuzzy rule-based node classification model is proposed for clustering. Its primary benefits are flexibility in selecting effective CHs, reliability in distributing CHs overload among the other nodes, and reducing communication overhead and cluster formation time in highly dense areas. In addition, we propose a routing scheme that balances the load among sensors. The proposed scheme is evaluated through simulations to compare our scheme with the existing algorithms available in the literature. The numerical results show the relevance and the improved efficiency of our scheme.

Keywords— Dense wireless sensor network; energy consumption; green clustering; network separation; fuzzy logic.

---

1. Introduction

In recent years, Wireless Sensor Networks (WSNs) have been used in a wide range of applications, such as medical treatments and tele healthcare, outer-space exploration, military applications, home applications and the monitoring of oceans, agriculture lands and wildlife [1-3]. In such applications, deployed sensor nodes are equated with limited on-board processing, storage and radio communication. Miniaturisation has enabled the fabrication of portable smart sensor nodes at low cost with high accuracy [4, 5]. Therefore, recently it has become commonplace to densely deploy a large number of sensor nodes within the target environment to increase the Quality of Service (QoS) of the network with respect to the data reception ratio and the robustness of the WSN.

*Corresponding author. Tel.:+91-9424346550, + 91 0751 2460313 (fax), E-mail: prasenjit.chanak@gmail.com (P. Chanak), ibanerjee@it.iiests.ac.in (I. Banerjee), r.s.sherratt@reading.ac.uk (R. S. Sherratt)

In such WSNs, each sensor node collects local information, processes the data and then transmits the collected information to the Base Station (BS) through single-hop or multi-hop communication. Distinguished from the traditional wireless networks, the deployed sensor nodes in WSNs are limited in power, computational capacity and memory. Battery replacement in WSNs may be impossible due to harsh and inaccessible deployment environments. According to energy consumption models [6, 7], energy consumption exponentially increases with required communication distance. Hence, transmission energy consumption is one of the most challenging issues for the long term operation of WSNs. Previous research has shown that multi-hop communication is more energy efficient than the single-hop communication [6, 8]. Therefore, clustering sensor nodes is an efficient strategy in terms of energy consumption where each deployed sensor node sends local information to a Cluster Head (CH) to which the node belongs [2, 9, 10] and CHs transmit their aggregated information to the BS [11, 12]. In clustered WSNs, CHs that are located close to the BS consume more energy than other CHs because they have to relay a relatively large number of packets, due to their position in relation to other CHs/nodes and the BS. Also, the CHs located close to BSs need to work as aggregators and they present higher computational burden than the rest CHs. As a result, CHs close to the BS tend to exhaust their energy (die) earlier compared to other CHs. In this circumstance, other CHs are unable to reach the BS causing the network to become disconnected, as presented in Fig. 1. However, most of the outer sensor nodes can still survive for a long period of time. This problem is commonly known as the premature death of the network [13-16] and is more crucial in the design of large scale dense WSNs. The premature death of the network can potentially reduce lifetime and QoS of the WSNs. If the load of the CHs and CM nodes can be reduced and balanced, then premature death of the network can be prevented and QoS of the network can be improved.

A number of clustering approaches have been proposed in the last few years where the CHs role is rotated among the non-CH nodes to balance energy consumption among the deployed sensor nodes [17-19]. However, rotation of CHs only balances energy consumption of the non-CH nodes in an intra-cluster environment, it does not balance the energy consumption among CHs in the inter-cluster multi-hop routing environment. Furthermore, most studies on load balanced clustering propose static solutions, that is, they do not adapt the topology to the changing network conditions [5, 20-24]. These conditions are often not favourable for the elected CHs, regarding energy consumption. As a result, the CHs frequently become non-operational, leading to new CHs-election, further increasing message overhead, and thus energy consumption in the network. Therefore, CH rotation mechanisms significantly increase message exchange among the deployed sensor nodes.

Fuzzy logic based clustering approaches [20-22] have been proposed to reduce the extra message exchange during the CHs selection phase. In these approaches, fuzzy logic has been used for the CH selection process, where higher energy level nodes are selected as CHs through the fuzzy rule-based process. These approaches save extra message overhead and time delay during the cluster formation phase. However, these approaches do not balance CH load in the inter-cluster environment. In order to address the balance of energy consumption problem in an inter-cluster environment, recent studies on balance energy consumption in WSNs have proposed unequal clustering mechanisms whereby the entire network was divided into several clusters of unequal size [23-24]. The main concept behind these unequal clustering mechanisms was to adjust the cluster size with respect to the distance between the CHs and the BS to reduce the energy consumption of the CHs that are close to the BS. It has been seen from the literature that the existing unequal clustering method [21, 23] leads to a large number of message exchanges over the network to determine the cluster size and for the exchange of current energy status information, which rapidly depletes energy of deployed sensor nodes and creates a large overhead for large scale dense WSNs. In addition, the unequal clustering method increases number of CHs within the network that creates huge traffic load within the network. Due to poor performance and high energy overhead of the existing set of approaches, it has become necessary to design and develop an efficient load balancing algorithm for large scale dense WSNs to prevent the premature death of the network and improve the QoS of the network.

In this paper, we propose a new energy aware fuzzy approach for large scale WSNs with different densities. We first present a fuzzy rule-based sensor node classification strategy that classifies deployed sensor nodes into different categories and selects a tentative CH set from the deployed sensor nodes. A fuzzy rule-based sensor node classification strategy helps to distribute the overload of the CHs and it is suitable for high uncertainty environments [25]. In order to reduce extra message overhead and time delay, we also propose a novel distributed CH selection algorithm that can select CHs from the tentative CH set. The proposed clustering algorithm adapts the topology to the changing network conditions, which increases the lifetime of the selected CHs and prevents frequent CHs selection process. Finally, we present a load balanced data routing algorithm that can minimize and balance load among the CHs.

The rest of the paper is organized as follows. Related works on the load balanced clustering in WSNs are presented in Section 2. Section 3 presents the network mode and explains the unbalance communication load and data routing problems. Section 4 describes proposed clustering and data routing algorithms. Complexity analysis and the energy consumption calculation of the proposed algorithms are parented in Section 5. In Section 6, we present simulation results which were conducted to measure the performance of the proposed algorithms. Finally, Section 7 concludes the paper.

2. Related Works

Recently, a significant amount of research has been done to address the energy consumption problem in WSNs. The main goals of these studies were to reduce the energy consumption among deployed sensor nodes and prolong the network lifetime.

Heinzelman et al. [26] proposed a distributed cluster-based data routing scheme called Low-Energy Adaptive Clustering Hierarchy (LEACH). LEACH has been a popular early study that uses a dynamic transmission range to balance energy consumption between the deployed sensor nodes. In this dynamic clustering approach, each deployed sensor node has a certain probability of becoming a CH per round and LEACH selects CHs based on the probability of the deployed sensor nodes. In order to balance energy consumption, LEACH dynamically rotates the work load of the CH among the other non-CH nodes. In LEACH, each CH communicates with the BS via single-hop communication which leads to much energy consumption. Hence, the LEACH protocol is not suitable and scalable for large scale networks. On the other hand, in this approach, a lower energy node may be selected as a CH due to probability based CHs selection process. If a lower remaining energy node is selected as a CH, then the CHs selection process is executed at small time intervals. To overcome this problem, Younis and Fahmy [27] presented a multi-hop data routing algorithm called Hybrid Energy-Efficient Distributed (HEED). In this approach, a two-phase parameter process was performed to select CHs. In the first phase, the remaining energy of the deployed sensor nodes is used for CHs selection. If a tie occurred in the first phase, the second phase parameters, such as node degree, intra-cluster energy consumption, and distance to neighbours, were considered to break the tie in the first phase CHs selection process. In this approach, CHs are well distributed over the monitoring field. However, HEED did not consider balancing energy consumption among CHs. Therefore, CHs close to the BS suffered from the high communication load and the CHs around the BS depleted their energy faster than the boundary CHs. As a result, HEED suffered from the premature death problem.

In order to overcome the uncertainties inherent in WSN environment, some cluster approaches have used fuzzy logic to overcome the problem. In the existing fuzzy based clustering approaches, CHs were selected by fuzzy rule-based mechanisms whereby fuzzy logic was employed to get a better combination of the input parameters to obtain an optimal output. CHEF was a fuzzy based CHs election algorithm where CHs election occurred in a distributed manner [20]. In this approach, the residual energy of each sensor node and local distance formed the fuzzy input parameters. CHs were elected on their residual energy and local distance of the deployed sensor nodes.

The main disadvantage of this approach was that the message overhead was very high compared to other cluster formation algorithms, and this approach did not consider the unequal load of the sensor nodes. Therefore, CHEF suffered from the network separation problem. Bagci and Yazici [21] proposed a fuzzy based unequal clustering approach where different unequal size clusters are designed to solve the network separation problem in a multi-hop WSN. The main drawback of this unequal clustering approach was that it increased the number of CHs within the network which significantly increased the traffic load through the deployed sensors. In addition, the unequal clustering approach leads to a large number of message exchanges over the network for status interchange during the unequal cluster formation phase which rapidly depleted the current energy of the deployed sensor nodes. Li et al. [28] proposed an unequal clustering scheme called Energy Efficient Unequal Clustering (EEUC) for balanced energy consumption among CHs. In this approach, unequal cluster size determination method led to a large number of message exchanges over the network. Therefore, this approach suffered from poor network lifetime. On the other hand, in this approach, it would be difficult to control the actual size of the clusters when the number of exhausted nodes is very high within the network. Sabor et al. [23] proposed an Unequal Multi-hop Balanced Immune Clustering protocol (UMBIC) to solve the network separation problem and improve the network lifetime. In this approach, unequal clustering and multi objective algorithm were used to balance intra-cluster and inter-cluster energy consumption. In addition, fuzzy logic was used to select CHs within the network. However, the main drawback of this approach was that the high volume of control message exchange between selected CHs and sensor nodes to from clusters within the network. Gajjar et al. [24] proposed an unequal clustering algorithm named Fuzzy and Ant Colony Optimization Based Combined MAC, Routing, and Unequal Clustering Cross-Layer Protocol for WSN (FAMACROW). In this technique, fuzzy logic was used to select CHs within the network. As a large number of message exchanges over the network consume more energy compared to information processing, this approach dissipated more energy and reduced the lifetime of the network. Furthermore, in this approach, data delivery delay was very high which is unsuitable for large scale dense WSNs. Rao and Banka [11] proposed an unequal clustering and routing algorithms for wireless sensor networks where a chemical reaction optimization based scheme was adopted for CH selection and multi-hop data routing. In this approach, unequal cluster size determination and multi-hop data routing path selection method leads to a large number of message exchanges over the network. It also required a long time for CH selection and multi hop data routing. Zahedi et al. [29] proposed a swarm intelligence based fuzzy routing protocol for clustered wireless sensor network, called Swarm Intelligence Fuzzy SIF. In this work, the residual energy, distance to the sink, and distance from the cluster were considered as fuzzy input parameters to select CHs within the network. This approach generated unequal traffic load among the CHs which significantly reduced the performance of the network. Semchedine et al. [30] proposed a load balanced algorithm for data centric routing in wireless sensor network called Directed Diffusion with Load Balancing (DDLB). In this approach, a routing protocol for energy efficient data transmission from the deployed sensor nodes to the BS was proposed. In addition, a load balanced algorithm was introduced for equal energy consumption of the sensor nodes. DDLB approach only balanced energy consumption among the deployed sensor nodes and CHs in intra-cluster environment. It did not balance energy consumption between the CHs in the inter-cluster environment. Therefore, it was unable to prevent the premature death of the WSNs. It also required a long time for data routing. Furthermore, it injected a large message overhead into large scale dense WSNs.

Above all, the researchers only considered intra-cluster load balancing or cluster head selection. In this paper, two factors are studied together. A routing algorithm for wireless sensor network is presented to balance energy and prolong the network lifetime.

3. Network models and problem statement

3.1 Network model

The following properties are assumed for the sensor network design [5, 11]:

- We assume a large scale dense WSN, where sensor nodes are powered by a non-renewable energy source. When this energy supply is exhausted, the sensor node becomes non-operational; otherwise sensor nodes sense data from monitoring environment. Nodes are also capable of data receiving and transmitting data.
- Initially, all sensor nodes are charged with the same amount of energy.
- Sensor nodes are non-uniformly or randomly distributed within the monitoring environment and all deployed sensor nodes are stationary.
- The BS is not limited in energy, memory, computational power.
- Deployed sensor nodes are not location aware.
- The WSN is homogeneous, all nodes have the same processing power, communication capability and memory capacity.
- The network stops working when all of the sensor node energy is exhausted.
- The distance between two deployed sensor nodes can be estimated using the receiving signal strength.
- Every deployed sensor node is capable of adjusting its transmission power, according to the distance of the destination nodes.

3.2 *Notations*

For clarity, we describe some notations that we have used throughout the paper in Table 1.

3.2 Energy model

In this work, the first order radio model is used for measuring the energy consumption of deployed sensor nodes. Energy consumption by the node (v_i) for single message transmission is represented by [10]:

$$E_T = \begin{cases} (\tau_t + \tau_{fs} d^2) P_i & if d < d_0 \\ (\tau_t + \tau_{amp} d^4) P_i & if d \geq d_0 \end{cases} \tag{1}$$

where E_T is the energy dissipated to transmit P_i bits data packet over a distance d and d_0 is a distance threshold where $d_0 = \sqrt{\tau_{amp}/\tau_{fs}}$. τ_t [J/bit] is the energy loss per bit by the transmitter circuitry, τ_{fs} (J/bit/m^2) and τ_{amp} (J/bit/m^4) denote the factors in the Friss' free space model. Energy dissipation by the receiver circuit for receiving P_i (bits) message is represented by:

$$E_R = \tau_r P_i \tag{2}$$

where E_R is the energy dissipated to receive P_i bits data packet. τ_r[J/bit] represents the energy dissipation by the receiver circuit. In addition, the energy consumption by a sensor node (v_i) for sensing P_i bits data is represented by:

$$E_{sen} = P_i E_{sensing} \tag{3}$$

In cluster based approaches, local information is highly correlated, and hence CHs use a data aggregation mechanism to combine correlated information into a single-fixed packet. The energy consumption for aggregation *m* packets of P_i bits is represented by:

$$E_{agg} = m P_i E_{DA} \tag{4}$$

where E_{DA} is the energy dissipated to calculate data correlation over mP_i bits data packet.

3.3 Problem statement

We consider a homogeneous WSN consisting of *N* sensor nodes. Sensor nodes are randomly deployed in a 2-dimensional plane. In the initial network topology, each sensor node has transmission range R_{max}. We represent the initial network topology with an undirected weighted graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots., v_N\}$ is the set of the deployed sensor nodes and $E = \{(v_i, v_j) | dist\ (v_i, v_j) < R_{max}\}$ is the set of edges, where $dist(v_i, v_j)$ depicts the distance between nodes v_i and v_j.

The network is divided into M number of clusters and each cluster contains $n_i(i = 1, 2, ..)$ member nodes. Each CM node senses P_i bits data from the surrounding region and transmits this data to which the node belongs. Therefore, the load of a CM node is given by:

$$E_{nonCH} = E_{sen} + E_T = P_i E_{sensing} + (\tau_t + \tau_{d1} {d_i}^2) P_i \tag{5}$$

where d_i is the average distance between a CM node and its CH. Furthermore, the energy consumption of the CHs primarily depends on sensing data, traffic load, and routing load. The traffic load in a CH is expressed as:

$$E_{trafficCH} = \frac{N}{M} \times (E_R) + (N_P) E_R = (n_i + N_P) \tau_r P_i \tag{6}$$

where N_p is the number of routing packets receive by a CH from other CHs. The routing load in a CH is given by:

$$E_{routingCH} = (\tau_t + \tau_{amp} {d_j}^4)(N_P + 1) P_i + (n_i + 1) P_i E_{DA} \tag{7}$$

where d_j is the average distance between a CH and the next hop CH or BS. The total acting load of a CH is expressed as:

$$E_{CH} = P_i E_{sensing} + \left(\left(\frac{N}{M} - 1\right) + N_P\right) \tau_r P_i + \left(\frac{N}{M}\right) P_i E_{DA} + (\tau_t + \tau_{amp} {d_j}^4)(N_P + 1) P_i \tag{8}$$

From eq. (8), it can be seen that the CH load is significantly more than a non-CH node within the cluster because a CH not only collects the data from its CM nodes, it also transmits the data to the BS. Hence, CHs close to the BS tend to exhaust their energy faster because of the heavy relay traffic. Therefore, the premature death of the network may arise and shorten the lifetime of the network. To solve this problem, some load balanced clustering approaches were proposed [23-24, 29]. However, existing load balanced clustering approaches put a large message overhead into the large scale WSNs which itself rapidly depletes the energy of the sensor nodes.

The energy consumption of a WSN depends on inter- and intra-cluster traffic load. The inter-cluster traffic depends on energy spent to communicate with other CHs as well as the BS. Intra-cluster traffic load depends on the energy consumption from data communication inside the cluster as well as data processing. In large scale dense WSNs, inter-cluster traffic load consumes more energy than the intra-cluster traffic load. Therefore, the premature death of the network is a core issue in the design of large scale dense WSNs. An energy efficient load balanced multi-hop routing algorithm is required to prevent the premature death of the WSNs, especially for large scale network. Thus objectives of this paper are (a) to reduce the energy consumption of the network, (b) reduce inter-cluster traffic load, and (c) balance energy consumption among the CHs and CM nodes. These three functions of a network can avoid the premature death of the WSNs and improve the lifetime of the network. Specifically, we present our objectives as follows:

Objective 1: Total energy consumption of the network (E_{total}) must be minimized. The total energy consumption of the network is given by:

$$E_{Total} = \sum_{j=1}^{M}\left[\left(P_i E_{sensing} + \left(\left(\frac{N}{M} - 1\right) + N_P\right)\tau_r P_i + \left(\frac{N}{M}\right)P_i E_{DA} + \left(\tau_t + \tau_{amp} {d_j}^4\right)(N_P + 1)P_i\right) + \sum_{i=1}^{\left(\frac{N}{M}-1\right)} P_i E_{sensing} + \left(\tau_t + \tau_{fs}{d_i}^2\right)P_i\right] + C_n\left(\sum_{i=1}^{N} k_i\left(\left(\tau_t + \tau_{fs}{d_i}^2\right) + \tau_r\right)\right) \quad (9)$$

where C_n is the number of control messages flowed over the network and k_i represents the size of the control messages. First part of eq. (9) represents the minimum energy consumption of the CH and the second part of eq. (9) describes the minimum energy consumption of the CH nodes. The remaining part of eq. (9) represents the energy loss of the deployed sensor nodes due to the required handling of control messages.

Objective 2: The number of control messages (C_n) must be minimized. C_n is given by:

$$\left(C_n = \left(\sum_{i=1}^{N} k_i\left(\left(\tau_t + \tau_{d1}{d_i}^2\right) + \tau_r\right)\right)\right) \quad (10)$$

Objective 3: The traffic load of the CHs ($E_{Traffic}$) should be minimised. $E_{Traffic}$ is given by:

$$\left(E_{Traffic} = \sum_{i=1}^{M}(n_i + N_P)\tau_r P_i\right) \quad (11)$$

Objective 4: The average load ($E_{Average\_load}$) between the CH and its CM nodes should be minimized. $E_{Average\_load}$ is:

$$E_{Average\_load} = \left(P_i E_{sensing} + \left(\left(\frac{N}{M} - 1\right) + N_P\right)\tau_r P_i + \left(\frac{N}{M}\right)P_i E_{DA} + \left(\tau_t + \tau_{d2}{d_j}^4\right)(N_P + 1)P_i\right) + C_n k_i\left(\left(\tau_t + \tau_{fs}{d_i}^2\right) + \tau_r\right) - M\left(\sum_{i=1}^{\left(\frac{N}{M}-1\right)} P_i E_{sensing} + \left(\tau_t + \tau_{d1}{d_i}^2\right)P_i + C_n k_i\left(\left(\tau_t + \tau_{fs}{d_i}^2\right) + \tau_r\right)\right)\Big/N \quad (12)$$

4. Design Rational and Proposed Scheme

To solve the balanced energy consumption problem in WSNs, previous research [20, 21] periodically rotated the role of the CH among the CM nodes. The rotation of the CHs can only balance the energy consumption among the CHs and CM nodes. It cannot balance energy consumption among the CHs in the inter-cluster environment. Therefore, our mechanism focuses on the energy balance among the CHs in the inter-cluster environment. In addition, previous clustering approaches [15, 22] potentially increase extra message overhead during the CHs rotation process which itself shortens the network lifetime. Therefore, this work also focuses on the message overhead problem. In this section, we first estimate the traffic load and energy consumption of the CM nodes, as well as CHs in intra-and inter cluster environment. After these, we propose fuzzy rule-base node classification strategy, clustering algorithm, and load balanced data routing algorithm. Fig. 2 shows a flow chart of our proposed scheme.

Theorem 1: Denote R_i as the time period for one data gathering round. d_i is the distance between node j and its CH. If node j processes C_n number of control messages at the R_i time period, the average energy consumption $E_{average}^{j}$ of j is $E_{average}^{j} = E_T^j + E_R^j + E_{sen}^j$, where:

$$\begin{cases} E_T^j = \left(\tau_t + x{d_i}^l\right)\left(\left(n_g + 1\right)P_i + C_n k_i\right) \\ E_R^j = \tau_r(C_n k_i + n_g P_i) \\ E_{sen}^j = P_i E_{sensing} \end{cases} \tag{13}$$

and if $d_i \leq d_0, x = \tau_{fs}$ and l=2; otherwise, $x = \tau_{amp}$ and l=4.

Proof: In a data gathering round, the energy consumption of node j consists of the following three parts.

1) Energy consumption for receiving data: a sensor node j receives n_g number of data packets from its neighbour nodes and C_n number of control messages in a round. Therefore, energy consumption for receiving is $E_R^j = \tau_r(C_n k_i + n_g P_i)$.
2) Energy consumption for data transmitting: a sensor node j transmits n_g number of data packets with its own sensing data. In addition, sensor node j also transmits C_n number of control messages in a round. Therefore, energy consumption for data transmitting is $E_T^j = \left(\tau_t + x{d_i}^l\right)\left(\left(n_g + 1\right)P_i + C_n k_i\right)$.
3) Energy consumption for data sensing: sensor node j senses P_i amount of data from its surrounding region at R_i time. Therefore, energy consumption for data sensing is $E_{sen}^j = P_i E_{sensing}$. In a round, the energy consumption $E_{average}^{j}$ of node j is $E_{average}^{j} = E_T^j + E_R^j + E_{sen}^j$.

Theorem 2: Assume that cluster i is in the region of B_x with the width of ϑ and B_x region is composed of N/M number of non-CH nodes. Denote d_I as the distance between CH i and the BS. If CH i process and relays C_n number of control messages and N/M number of data packets at the R_i time period, the average energy consumption $E_{average\_CH}^{i}$ of i is $E_{average\_CH}^{i} = E_T^i + E_R^i + E_{sen}^i + E_{DA}^i$, where:

$$\begin{cases} E_T^i = \left(\tau_t + x{d_I}^l\right)\left((N_P + 1)P_i + k_i C_n\right) \\ E_R^i = \left(\left(\left(\dfrac{N}{M} - 1\right) + N_P\right)P_i + k_i C_n\right)\tau_r \\ E_{DA}^i = \left(\dfrac{N}{M}\right)P_i E_{DA} \\ E_{sen}^i = P_i E_{sensing} \end{cases} \tag{14}$$

and if $d_I \leq d_0, x = \tau_{fs}$ and l=2; otherwise, $x = \tau_{amp}$ and l=4.

Proof: In a data round, the energy consumption of CH i consists of the following four parts.

1) Energy consumption for data receiving: CH i receives all data packets from its N/M number of CM nodes and C_n number of control messages in a round R_i. In addition, CH i relays data packets from N_p number of other CHs. Therefore, the energy consumption in receiving data is:

$E_R^i = \left(\left(\left(\frac{N}{M} - 1\right) + N_P\right) P_i + k_i C_n\right) \tau_r.$

2) Energy consumption for transmitting data: in a round, a sensor node i transmits aggregated data packets to the BS. In addition, CH i also exchanges C_n number of control messages with its CM nodes. CH i relays N_p number of data packets from other CHs. Therefore, energy consumption in transmitting data is $E_T^i = (\tau_t + x d_I{}^l)((N_P + 1)P_i + k_i C_n)$.

3) Energy consumption for data aggregation: CH i aggregates correlated data into a single-fixed packet. Therefore, energy consumption for data aggregation is $E_{DA}^i = \left(\frac{N}{M}\right) P_i E_{DA}$.

4) Energy consumption for sensing: CH i senses P_i amount of data in a round. Therefore, energy consumption for data sensing is $E_{sen}^i = P_i E_{sensing}$. In a round, energy consumption $E_{average\_CH}^i$ of CH i is $E_{average\_CH}^i = E_T^i + E_R^i + E_{sen}^i + E_{DA}^i$.

According to the theorem 1 and theorem 2, several phenomena can be concluded as follows:

1) Traffic load and energy consumption have a direct relationship with transmission radius R_{max}, which may cause the location of the network separation deviating from the adjacent area of the BS.
2) When R_{max} is fixed, the total energy consumption is impacted by the energy consumption for the transmission and reception of data packets.

Since the first energy-exhausted nodes must be the ones with the maximum energy consumption in the network, the time duration of the first node being exhausted within the network is:

$$R_L = \left\lfloor \frac{E_{max}}{\max\left(E^i_{average_{CH}}\right)} \right\rfloor \tag{15}$$

where E_{max} is the maximum energy of the sensor nodes. If the time period of the first exhausted node within the network can be extended, then the premature death of the network can be faced. In the next section, we propose a load balanced data routing scheme where the overload of the CHs is reduced through the routing nodes. CHs overload is computed using eq. (14). Control message overhead also increases energy consumption of the CM nodes, as well as CHs. Therefore, in this paper, we also reduce control messages flow within the network.

Sensor nodes can typically be deployed in very dynamic changing environments, where network variables frequently change their values. Energy status, inter-node distance, acting load within a node, and size of the network are a few of the variables which need to be continually monitored. In such a case, fuzzy rule-based approaches are very effective to overcome the uncertainties inherent in the WSN environment. The total energy consumption of the network will depend on the number of control message flows over the network (eq. (13) and eq. (14)). If the number of control messages is decreased, then the total energy consumption of the network also decreases. Therefore, in this paper, we reduce the number of control messages within the network through the fuzzy-rule based nodes classification process. The proposed fuzzy rule-based nodes classification strategy classifies deployed sensor nodes into different categories and defines a tentative CH set. Therefore, our proposed clustering algorithm selects CHs from the tentative CHs set through the minimum number of nodes involvement to significantly reduce the number of control message flows within the network. The proposed scheme is divided into three phases, viz (a) nodes classification through fuzzy logic, (b) cluster formation phase, and (c) load balanced data routing phase. The operation of the proposed scheme is split into rounds, where each round consists of three phases as shown in Fig. 3. The first phase of the proposed scheme is the node classification phase with T_1 duration. In node classification phase, deployed sensor nodes are classified into three categories: (i) strong node S_g, (ii) moderate node M_d, and iii) weak node W_k with the help of fuzzy rule-based. This phase is distributed, where each deployed sensor node makes own decision through the fuzzy inference system. The second phase of the proposed scheme is a cluster formation phase with T_2 duration. In this phase, the whole network divided into several clusters. The best CHs searching process among all the deployed sensor nodes is a complicated process and it needs a long time especially in large scale dense WSNs. To solve this problem, in our proposed scheme, only strong nodes participate in the CHs selection process. The third phase of the proposed scheme is a load balancing and routing phase with T_3 duration. The data routing phase is divided into two sub-phases where each CH calculates its acting load and selects a routing node through the distributed manner. On the other hand, after the routing node selection phase, CM nodes transmit their data to a respective CH and the CH relays aggregated data to the BS directly, or via a routing node.

4.1 Nodes classification through fuzzy logic inference system

A Fuzzy rule-based system has been used for node classification, where deployed sensor nodes have been classified into three different categories (strong, moderate, weak) according to their routing load, traffic load, and current energy level condition. The strong node set V_g ($V_g \subseteq$V) has minimum load and maximum residual energy. Therefore, CHs are primarily selected from a strong node set (V_g). The moderate node set V_d ($V_d \subseteq$V) has medium load and medium residual energy. Hence, V_d node set senses data from monitoring environment and transmits to the adjacent CH. The moderate node takes the supplementary role in managing the cluster once the strong node's energy starts to drain out erratically. The weak node set V_k ($V_k \subseteq V$) has maximum load and minimum residual energy. So, weak node acts as a cluster member node. The week node does not perform any aggregation or communication with the BS and this makes the node retain its energy for sensing data. After classification of the deployed sensor nodes, strong node set defines as tentative CHs. The proposed clustering algorithm selects final CH set from the tentative CHs without any central control. The fuzzy logic system is divided into four parts, (a) fuzzifier, (b) inference engine, (c) fuzzy rule base, and (d) defuzzifier. In the fuzzifier module, input variables are transferred to a linguistic value according to the membership function. The output of the fuzzifier is the input of the fuzzy inference engine. Fuzzy rule-base is a set of linguistic control rules. The inference engine makes decisions based on the fuzzy control rule and input of linguistic variables. Defuzzifier generates a non-fuzzy control output from the inferred fuzzy control action.

In order to classify deployed sensor nodes, three input variables are used. The first one is the communication load or routing load. Each node calculates its communication load based on its distance from the BS. The second fuzzy input variable is the traffic load of the sensor node. The third input variable is the residual energy of the sensor node. Three linguistic variables "Low", "Medium", and "High" are used for communication load. The linguistic variables for traffic load are "Low", "Medium", and "High". Similarly, linguistic variables for residual energy are "Low" "Medium" and "High". The output of the inference engine refers to the appropriateness of the node as a strong node, moderate node, weak node, and poor node (Fig.4). Poor nodes are energy exhausted nodes. Therefore, these nodes are excluded from the network and are not considered further. Table 2 lists the rule-base employed by the inference engine. A Mamdani Controller [31] is used as a fuzzy inference technique to evaluate the rule set and the Centre of Area (COA) method is used for defuzzification. Mamdani controller consists of: (1) combining the fuzzified inputs according to the fuzzy rules in order to establish rule strength, (2) finding the consequence of the rule by combining the rule strength and the output membership function, and (3) combining the consequences to generate an output distribution. After classifying deployed sensor nodes using Mamdani controller, the cluster formation algorithm is used to select CH based on residual energy and acting load respectively. According to the fuzzy rules, higher residual energy and lower acting load nodes are selected as tentative CHs. The proposed clustering scheme selects CHs from the tentative CH set depending on their current energy condition which is described in the next section.

4.2 Cluster formation

After classification of the deployed sensor nodes, sensor nodes are organized into different clusters. In the proposed scheme, only higher residual energy and lower acting load sensor nodes are participated in CHs selection process without any central control (Fig.4 and Fig. 5). In the clustering phase, each strong sensor node ($v_g \subseteq V_g$) sets its own timer independently before starts CH advertisement (*ADVE* message). Let T_{s_i} be the timer of a strong node v_g^i which is derived as:

$$T(v_g^i) = \frac{E_{max}(v_g^i) - E_{current}(v_g^i)}{E_{max}(v_g^i)} \times T_{CH} \tag{16}$$

where T_{CH} is the maximum allocated time for CH selection and E_{max} is the initial maximum energy of the strong node v_g^i. $E_{current}$ is the residual energy of the strong node v_g^i. According to the eq. (16) a strong sensor node with higher residual energy, lower communication load, and traffic load will be selected as CH from the strong node set (V_g). Once the timer expires then the node v_g^i selects itself as a CH and broadcasts *CH_ANNOUNCE* message in the communication range R_{max}. The *CH_ANNOUNCE* message includes Frame Identification (FI), Source Identification number (SID), Residual Energy (RE), and location information. Fig. 6 shows the *CH_ANNOUNCE* message format. If a strong node v_g^j receives the *CH_ANNOUNCE* message, then it withdraws its nomination for CH by cancelling its timer and it acts as a non-CH node for the upcoming data routing round. Strong node v_g^j also starts keeping track of the sensor nodes from which it receives a *CH_ANNOUNCE* message by maintaining a neighbour CH set denoted by $N_{CH}(v_j)$. Similarly, when moderate nodes (V_d), and weak nodes (V_k) receive *CH_ANNOUNCE* message then they also keep track of sensor nodes from which they receive a *CH_ANNOUNCE* message by maintaining a neighbour CH set ($N_{CH}(v_j)$). All non-CH nodes decide their cluster membership later by using $N_{CH}(i)$.

To form the clusters, each non-CH node decides its cluster membership as follows. A sensor node needs to join one of the CHs belonging to the set $N_{CH}(v_j)$. Let $v_g^1, v_g^2, v_g^3, \ldots, v_g^m$ be the set of CHs belonging to the set $N_{CH}(v_j)$. Non-CH node v_j ($v_j \in (V_g \cup V_d \cup V_k)$) computes the average communication load of CHs, denoted by $L_{avg}(v_j)$ which is calculated as follows

$$L_{avg}(v_j) = \frac{\sum_{i=1}^{m} E(n_i, P_i, d_i)}{m} \tag{17}$$

Non-CH node v_j joins the nearest CH whose communication load is less than or equal to $L_{avg}(v_j)$ by broadcasting a cluster joint message *CH_JOIN* in the communication range R_{max}. The detail description about this algorithm is summarized in algorithm 1.

Algorithm 1: cluster formation

/* cluster formation*/

Initial, $\forall V$ classified into $(V_g \cup V_d \cup V_k)$.
If $v_g^i \subseteq V_g$ **then**
    Set timer according to eq. no. 16.
    Node v_g^i broadcast *ADVE* message.
end if
if (v_g^j receive *CH_ANNOUNCE* msg.) **then**
    Node v_g^j stop *CH_ANNOUNCE* by switches off its timer.
    v_g^j sets as a non-CH node.
    v_g^j Updates $N_{CH}(v_j)$.
end if
for V
    if $v_i \subseteq (V_d \cup V_k)$ then
        v_i sets as a non-CH node.
        v_i update $N_{CH}(v_i)$.
    end if
end for
for each non-CH node $v_m \subseteq (V_g \cup V_d \cup V_k)$.
    sum=0.0;
    for each CH v_g^k belong to $N_{CH}(v_i)$
        $sum = sum + E_{CH}(v_g^k)$;
    end for
    $L_{avg}(v_j) = sum / |N_{CH}(v_i)|$
    for each CH v_g^k belongs to $N_{CH}(v_i)$
        if ($E_{CH}(v_g^k) \geqslant L_{avg}(v_j)$)
            $A_{CH} = v_g^k$;
        end if
    end for

28. Join with A_{CH} CH.

29. **end for**

4.3 Routing nodes selection and data routing phase

To route the data to the BS and balance energy consumption of CHs, CHs select routing nodes depending on their acting load. In this phase, routing nodes are selected without any central control. In order to avoid the dissipated energy in the overhead control packet and high computational time of the data routing phase, the proposed scheme performs routing node selection phase only if the acting load of any selected CH greater than the average load of the cluster (AC_{Load}). Initially, each CH computes the average load of the cluster AC_{Load} is

$$AC_{Load} = \frac{M\ (E_{CH} + \sum_{i=1}^{\left(\frac{N}{M}-1\right)} E_{nonCH})}{N} \tag{18}$$

If a CH v_g^i detects its acting load is greater than the AC_{Load}, it broadcasts a *TR_REQ* message in the range R_{max}. The message contains frame identification, source ID, residual energy ($E_{current}$ (v_j)), its distance from the BS (D_{CH}), and location information. Fig. 7 shows the *CH_ANNOUNCE* message format. If a sensor node v_i receives *TR_REQ* message and its distance from the BS (D_{v_i}) is less than the D_{CH} and residual energy is greater than the v_g^i, it sends a *TR_REP* message to request CH v_g^i. The message contains its ID, D_{v_i}, current energy level ($E_{current}$ (v_i)), and location information. When v_g^i receives the *TR_REP* message, then it is set as one of the routing nodes by maintaining a neighbour routing node set denoted by *TN(*v_g^i*)*. Otherwise, it simply discards the message. Recursively, all CHs whose acting load is higher than the average load of the cluster broadcasts the *TR_REQ* message to complete the routing node selection process.

In the routing node selection process, a CH may have multiple *TNs(*v_g^i*)* and hence multiple path to the BS. Let $v_1, v_2, v_3, \dots, v_k, \dots \dots, v_p$ are the set of routing nodes belonging to routing node set *TNs(*v_g^i*)*. Each CH v_g^i computes the average residual energy of the routing nodes, which is calculated as:

$$\delta_{avg}(v_R) = \frac{\sum_{i=1}^{p} E_{current}(v_i)}{p} \tag{19}$$

where $\delta_{avg}(v_R)$ is the average residual energy of the *TNs(*v_g^i*)* node set. CH v_g^i selects a routing node for data routing from the *TN(*v_g^i*)* whose residual energy is greater than or equal to $\delta_{avg}(v_R)$ and communication load is less than the communication load of the CH v_g^i . Similarly, each selected routing node v_R checks its traffic load $T_{Load}(v_R)$. If $T_{Load}(R)$greater than the communication load of v_R, the v_R selects another routing node in a similar way. The detailed description about this algorithm is summarized in algorithm 2.

Algorithm 2: Load balanced data routing algorithm

/* Routing node selection*/

for each CH v_g^i

CH calculates communication load $E_{routingCH}$.

CH calculates average communication load AC_{Load} of the cluster by eq. no. 14.

end for

for each CH v_g^i

if $E_{CH} > AC_{Load}$ **then**

CH v_g^i broadcasts *TR_REQ* msg. in the R_{max} rang.

end if

end for

for each node v_i

if $D_{v_i} \leq D_{CH}$ **then**

Node v_i sends *TR_REP* msg.

end if

end for

for each CH v_g^i

sum=0.0;

for each node v_i belong to $TNs(v_g^i)$

sum=sum+$E_{current}(v_i)$;

end for

$\delta_{avg}(v_i) = sum/\left|TNs(v_g^i)\right|$;

for each node v_i belong to $TNs(v_g^i)$

if $E_{current}(v_i) \geq \delta_{avg}(v_R)$ and $E_{CH}(j) \leq E_{CH}(R)$ **then**

Node v_i selects as routing node;

end if

end for

/*Load balancing of the selected routing nodes*/

for each routing node v_k

if $E_{CH}(R) > T_{Load}(v_R)$ **then**

28. Recursively, step 5 to 25 continues for selecting another routing node.

29. **end if**

30. **end for**

---

5. Complexity and Discussion

In this section, we present the analysis of message complexity, time complexity, and estimate the network lifetime of the proposed algorithms.

Lemma 1: *Message and time complexity of the proposed clustering algorithm is O(1) per sensor node and O (N) for N sensor nodes in the network.*

Proof: In the proposed clustering algorithm, a sensor node v_i ($v_i \in (V_g \cup V_d \cup V_k)$ either broadcasts *ADVE* message or *CH_JOIN* message only. Therefore, the message complexity of the proposed clustering algorithm is *O (1)*. On the other hand, in the proposed clustering algorithm, each sensor node decides independently whether to become a CH or not, which can be done in constant time. For cluster formation, each sensor node needs to process *N-1* messages in worst case to join a cluster. Therefore, the time complexity of the proposed clustering algorithm is *O(N)*.

Lemma 2: *The time complexity of the proposed data routing algorithm is O (N) for N sensor nodes in the network.*

Proof: In the proposed data routing algorithm, each CH needs to calculate the average residual energy of deployed sensor nodes for which it requires to check the residual energy of *N-1* nodes in the worst case. Therefore, the time complexity of the proposed routing algorithm is *O(N)*.

Lemma 3: *The network lifetime (NLT) due to the proposed algorithm is min* $\{TE_{max}/EL_i\}$ *i=1, 2, 3….N where* TE_{max} *and* EL_i *are the total energy present at deployed sensor nodes and consumed energy by the deployed sensor nodes respectively.*

Proof: The network lifetime (*NLT*) is defined as the total number of data gathering round upto last node die within the network due to the energy consumption. Let TE_{max} be the total initial energy of the network. A deployed sensor node v_i utilizes PE_i amount of energy for data processing, the CTE_i amount of energy of cluster formation and routing node selection, the EOA_i amount of energy for other activity of the network. Therefore, the network lifetime due to the proposed algorithms is defined as $NLT = \min\{TE_{max}/EL_i\}$ where $EL_i = PE_i + CTE_i + EOA_i$.

6. Results

In this section, we compare performance of the proposed scheme with the recent existing fuzzy based routing algorithms UMBIC [23], FAMACROW [24], and SIF [29]. The performance of the proposed scheme has been obtained by extensive simulations using the NS3 tool. We perform our simulations in various scenarios where a large number of sensor nodes are deployed in a square area. All the deployed nodes are homogeneous and initial energy of all the nodes has been set to 0.5J. All deployed sensor nodes generate information with a sensing rate of 200 bps. The detailed simulation parameters with their values are given in Table 3 [23, 24]. Simulations are conducted in two scenarios and the detail description of the simulation scenarios are as follows.

1. **Scenario 1 (low density networks):** 100 to 500 sensor nodes are randomly distributed in a field with dimension 100×100 [m^2]. Without loss of generality, a BS is placed at the centre of the monitoring region, i.e., (50, 50). Fig. 8.a shows a low density network scenario.
2. **Scenario 2 (dense networks):** 500 to 1000 sensor nodes are randomly deployed in a field with dimension 100×100 [m^2]. A BS is located at the outside of the network, i.e., (250, 200). Fig. 8.b shows a large scale dense network scenario.

6.1 Total energy consumption

In Fig. 9, we compare the total energy consumption of the proposed scheme, UMBIC, SIF, and FAMACROW. We can see that proposed scheme minimizes the total energy consumption of all deployed sensor nodes in the WSN that increases the overall lifetime of the network. Fig. 9.a shows the energy consumption of the network in scenario 1. It demonstrates that the proposed scheme is up to 37% more energy-efficient compared to UMBIC, 35% more energy-efficient compared to SFI, and 32% more energy-efficient compared to FAMACROW. Further, Fig. 9.b shows the energy consumption of the network in scenario 2. It shows that the proposed scheme is up to 27% more energy-efficient compared to UMBIC in dense network, 25% more energy-efficient compared to SIF, and 23% more energy efficient compared to the FAMACROW. From Fig. 9.a and 9.b, we can observe that if the node density is increased within the network, energy consumption rates also increases. This is because in large scale dense network, traffic load of the CHs is high compared to the low density networks. However, our proposed scheme achieved better results in terms of energy consumption in large scale dense network due to effective traffic management by the traffic node. In addition, according to theorem 1, the message overhead also has increased the energy consumption rate of all sensor nodes in the intra-cluster environment. In our proposed scheme, the message overhead is significantly less compared to other approaches due to our fuzzy rule-based node management strategy that significantly reduces energy consumption in our proposed scheme, especially in a dense network scenario.

6.2 Energy balancing

In this section, we examine the energy balance among CHs. Fig. 10.a and Fig. 10.b show the amount of the average energy consumption of CHs in a low density network (Scenario 1) as well as high density network (Scenario 2). As seen in the Fig. 10.a and Fig. 10.b, the average energy consumption of CHs is nearly equal in our proposed scheme. This is because the proposed scheme reduces inter-cluster communication load through the selection of efficient routing nodes that reduces and balances the energy consumption among CHs as equally as possible.

6.3 Alive nodes over rounds

Fig. 11.a shows that alive nodes within the network in scenario 1. In Fig. 11.a, it is seen that after 700 rounds the proposed scheme contains more than 61% active nodes as compared with UMBIC, 57% active nodes as compared with SIF, 55% active nodes as compared with FAMACROW in the low density network scenario. This is an indication of improvement in the performance of the network. It is due to the fact that the proposed scheme reduces energy consumption of the deployed sensor nodes through the effective load management proposes and preserves node lifetime. On the other hand, as we can see from the Fig. 11.b, the network lifetime of the proposed scheme is maximized as compared to other approaches in the dense network scenario. In dense network, after 1350 rounds proposed scheme contains more than 64% active nodes as compared with UMBIC, 58% active nodes as compared with SIF, 56% active nodes as compared with FAMACROW. We can find that proposed scheme is more effective for dense WSNs. This is because the proposed scheme effectively distributes high traffic load of CHs in dense network scenario through the routing nodes selection process that protects nodes death in the highly dense scenario. On the other hand, UMBIC, FAMACROW, and SIF reduce energy consumption of the CHs through next hop CHs selection process. It can potentially increase relay traffic of the CHs which are near to the BS. Therefore, in UMBIC, FAMACROW, and SIF, nodes die very fast compared to the proposed scheme.

6.4 Network lifetime

Fig. 12 shows the comparison in terms of First Node Exhausted (FNE), Half of Nodes Alive (HNA), and Last Node Exhausted (LNE) at scenario 1. The proposed scheme improves the stability period of FNE by 190, 160, 120 rounds as compared to UMBIC, SIF, and FAMACROW. On the other hand, the performance of the proposed scheme in terms of HNA is 150, 120, 80 rounds better as compared to UMBIC, SIF, and FAMACROW. Similarly, if LNE is considered, the performance of the proposed scheme is 170, 180, 110 rounds better than UMBIC, SIF, FAMACROW. It is due to effective load management between the CHs, as well as non-CH nodes. Fig. 13 shows the comparison results in term of FNE, HNA, and LNE at dense network (scenario 2). As seen in the Fig. 13, the proposed scheme also performs better than the UMBIC, SIF, and FAMACROW in dense network. It is due to the fact that the proposed scheme reduces energy consumption of all deployed sensor nodes in the intra-cluster as well as an inter-cluster environment that can significantly improve the lifetime of the deployed sensor nodes in highly dense scenario.

6.5 Scalability analysis

Fig. 14.a and Fig. 14.b show the average energy consumption in scenario 1 and scenario 2. This is the measure of the ratio between the sums of energy consumption of all deployed sensor nodes to the total number of deployed sensor nodes. In low density network (scenario 1), the average energy consumption rate of the proposed scheme is less by 37.5% as compared to UMBIC, 35% as compared to SIF, and 28.6 % as compared to FAMACROW. On the other hand, in high density network (scenario 2), the average energy consumption rate of the proposed scheme is less by approximately 43.4 % as compared to UMBIC, 42% as compared to SIF, and 36% as compared to FAMACROW. This is caused due to the elimination of load in inter-cluster data routing environment. In addition, our proposed scheme also takes advantage in fuzzy rule-based node calcification strategy that can potentially reduce extra overhead.

Fig. 15.a and Fig. 15.b show that average data transmission latency in scenario 1 and scenario 2. The average data transmission latency is least in the proposed scheme compared to the other scheme. This is because the proposed scheme reduces network traffic through the routing nodes selection process. However, UMBIC, SIF, FAMACROW select next hop CH to transmit their data to the BS. It increases traffic delay during the data transmission process.

Fig. 16.a and Fig. 16.b show the comparison results in terms of the data delivery ratio at scenario 1 and scenario 2. As can be seen, the proposed scheme performs much better than UMBIC, SIF, FAMACROW in a low density network scenario, as well as a high density network scenario. This is mainly because less data congestion in the proposed data routing scheme. In the proposed data routing scheme, CHs select routing node based on its traffic load. Traffic node mainly reduces traffic congestions in the CHs. Therefore, packet loss in the proposed scheme is less compared to the other schemes.

7. Conclusions

In this paper, a new energy aware fuzzy approach has been introduced for large scale Wireless Sensor Networks (WSNs) to prevent the premature death of the network. Our proposed scheme classifies deployed nodes into three different categories through a fuzzy logic process which significantly reduces the extra message overhead within the network and saves Cluster Head (CH) selection time, especially in the highly dense scenarios. It also helps to distribute overload of the CHs. The proposed scheme contains a novel distributed clustering algorithm where CHs are preserved for a sufficient amount of time. In order to avoid the premature death of the network which appears in large scale dense WSNs, we propose a new data routing algorithm that reduces and balances energy consumption of the CHs in inter-cluster multi-hop environment. The proposed routing algorithm computes the acting load of each CH and the overload of the CH is reduced through the routing node. Our scheme extends node lifetime and significantly avoids the premature death of the network. The results show that the proposed fuzzy-rule based approach achieves excellent performance in terms of average energy consumption, network lifetime, first node exhausted, half of nodes alive, last node exhausted, and energy variation of the CHs.

In the future works, we will try to improve the performance of our proposed scheme by applying other intelligent techniques derived from nature inspired algorithms.

Acknowledgments

This work is supported by SEED Division project under the Department of Science and Technology (DST) Govt. of India (SP/YO/036/2017(C)).

REFERENCES

[1] Cheng, H., Su, Z., Xiong, N., Xiao, N.: 'Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model', *Information Sciences*, 2016, 329, (1), pp. 461-477

[2] Chanak, P., Banerjee, I., Sherratt, R. S.: 'Energy-aware distributed routing algorithm to tolerate network failure in wireless sensor networks', *Ad-Hoc Network, 2017, 56, (3)*, pp. 158-172.

[3] Arivudainambi, D., Rekha, D.: 'Memetic algorithm for minimum energy broadcast problem in wireless ad hoc networks' *Swarm and Evolutionary Computation*, 2013, 12, (1), pp. 57–64

[4] Nguyen, T. D., Khan, J. Y., Ngo, D. T.: 'Energy harvested roadside IEEE 802.15.4 wireless sensor networks for IoT applications', *Ad Hoc Networks*, 2017, 56, (1), pp. 109–121

[5] Hoang, C. D., Yadav, P., Kumar, R., Panda, S.,: 'Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks', *IEEE Trans. Industrial Informatics.* 2014, 10, (1), pp. 774-783

[6] Shivappa, N., and Manvi, S.: 'Fuzzy inference system based 3D geographic routing in wireless sensor networks', *IET Wireless Sensor Systems*, 2019, 9,(3), pp. 132 –142

[7] Shankar A., Jaisankar, N., Khan, S. M., Patan, R., Balamurugan, B.: 'Hybrid model for security-aware cluster head selection in wireless sensor networks', IET Wireless Sensor Systems, 2019, 9, (2), pp. 68 –76

[8] Chanak, P., Banerjee, I., Samanta, T., Rahaman, H.: 'FFMS: Fuzzy Based Fault Management Scheme in Wireless Sensor Network'. In: *Proceedings of International Conference on Advances in Computing and Communications*. *Kerala India*, 2012, pp. 30-38

[9] Ghosh, N., Banerjee, I., Sherratt, R. S.: 'On-demand fuzzy clustering and ant-colony optimisation based mobile data collection in wireless sensor network', *Wireless Networks*, 2017, *23, (149)*, pp. 1635-6

[10] Rao, P. C. Srinivasa, Banka, H., Jana, Prasanta K.: 'Energy Efficient Clustering for Wireless Sensor Networks: A Gravitational Search Algorithm'. In: *Proceedings of International Conference on Swarm, Evolutionary, and Memetic Computing*.2015, pp 247-259

[11] Jayekumar, M. Nagarajan, V.: 'A novel DEA-OR algorithm for route failure recovery in dense wireless sensor networks', *Cluster Computing*.2018, 20(78), pp. 1-9

[12] Rao, P. C. Srinivasa, Banka, H.: 'Energy efficient clustering algorithms for wireless sensor networks: novel chemical reaction optimization approach', *Wireless Networks*. 2017, 23(2), pp. 433–452

[13] Anastasi, G., Conti, M., Francesco, M. D., Passarella, A.: 'Energy conservation in wireless sensor networks: A survey', *Ad Hoc Netw.,* 2009, 7, (3), pp. 537-568

[14] Ehsan, S., Hamdaoui, B.: 'A survey on energy –efficient routing techniques with QoS assurances for wireless multimedia sensor networks', *IEEE Communication Surveys & Tutorials,* 2012, 14, (2), pp. 265-278

[15] Li, F., Luo, J., Wang, W., and He, Y.: 'Autonomous deployment for load balancing k-surface coverage in sensor network', *IEEE Trans. Wireless Communications*, 2015, 14, (1), pp. 279-293

[16] Chanak, P., Samanta, T., Banerjee, I.: 'Cluster head load distribution scheme for wireless sensor networks'. In: *Proceedings of IEEE 2013 Sensor Conference, Maryland, USA*, 2013, pp. 1-4

[17] Meng, M., Wu, X., Xu, H., Jeong, B.S., Lee, S., Lee, Y.-K. : 'Energy efficient routing in multiple sink sensor networks'. In: *Proceedings of Fifth International Conference on Computational Science and its Application,* 2015, pp. 561-566

[18] Hwang, S.F., Lin, H., Cow, R. C.: 'An energy-efficient routing protocol in wireless sensor networks with holes'. In: *Proceedings of Fourth International Conference on Ubiquitous and Future Networks,*2012, pp. 17-22

[19] Li, C., Bai, J., Gu, J.,Yan, X., Luo, Y.: 'Clustering routing based on mixed integer programming for heterogeneous wireless sensor networks', *Ad Hoc Networks,* 2018, 72, (1), pp. 81-90

[20] Kim, J. M., Park, J. M., Han, S., Chung, Y. T.: 'CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks'. *In: Proceeding of International Conference of Advanced Communication Technology,*2008, pp. 654-659

[21] Bagci, H. Yazici, A.: 'An energy aware fuzzy approach to unequal clustering in wireless sensor networks', *Applied Soft Computing,* 2013, 13, (4),pp. 1741-1749

[22] Taheri, H., Neamatollahi, P., Younis, O.M., Naghibzadeh, S., Yaghmaee, M.H.: 'An energy – aware distributed clustering protocol in wireless sensor networks using fuzzy logic', *Ad Hoc Networks.*2012, 10, (7),pp. 1469-1481

[23] Sabor, N., Zahhad, M. A., Sasaki, S., Ahmed, S. M.: 'An unequal multi-hop balanced immune clustering protocol for wireless sensor networks', *Applied Soft Computing.*2016, 43, (2),pp. 372–389

[24] Gajjar, S., Sarkar, M., Dasgupta, K.: 'FAMACROW: Fuzzy and ant colony optimization based combined mac, routing, and unequal clustering cross-layer protocol for wireless sensor networks', *Applied Soft Computing,* 2016, 43, (2),pp. 235–247

[25] Mamdani, E.H.: 'Application of fuzzy algorithms for control of simple dynamic plant'. In: *Proceedings of Electrical Engineers*, 1974, pp. 1585–1588

[26] Heinzelman, B.W., Chandrakasan, P.A., Balakrishnan, H.: 'An Application-specific protocol architecture for wireless microsensor networks', *IEEE Trans. Wireless Communications.*2002, 1 , (4), pp. 660-670

[27] Younis, O., Fahmy, S.: 'HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor network', *IEEE Trans. Mobile Computing,* 2004, 3 (4), pp. 366-379

[28] Li, C., Ye, M., Chen, G., Wu, J.: 'An energy-efficient unequal clustering mechanism for wireless sensor networks'. In: *Proceedings of IEEE International Mobile Ad hoc and Sensor Systems Conference*,2005, pp. 1-8

[29] Zahedi, Z. M., Akbari, R., Shokouhifar, M., Safaei, F., Jalali, A.: 'Swarm intelligence based fuzzy routing protocol for clustered wireless sensor networks', *Expert Systems with Applications,* 2016, 55, (1), pp. 313–328

[30] Semchedine, F., Medjkoune, B. L., Tamert, M., Mahfoud, F., Aissani, D.: 'Load balancing mechanism for data-centric routing in wireless sensor network', *Computers and Electrical Engineering,* 2015, 41, (1), pp. 395–406

[31] Zimmermann, J. H. 'Fuzzy set theory and its applications'. *Fourth Edition, Kluwer Academic Publishers Group* 2001

Table 1
Notations used in the problem formulation and proposed algorithms

Symbol	Description
$\boldsymbol{V}$	The set of deployed sensor nodes, i.e., $V = \{v_1, v_2, \ldots\ldots v_N\}$
$\boldsymbol{E}$	The set of edges, i.e., $E = \{(v_i,v_j) \mid dist(v_i,v_j) < R_{max}\}$
$\boldsymbol{R_{max}}$	The maximum transmission range of a sensor node
$\boldsymbol{M}$	Number of clusters
$\boldsymbol{n_i}$	Number of cluster member nodes within a cluster
$\boldsymbol{d_i}$	The average distance between a cluster member node and its CH
$\boldsymbol{N_p}$	Number of routing packets receive by a CH from other CHs
$\boldsymbol{d_j}$	The average distance between a CH and the next hop CH
$\boldsymbol{C_n}$	Number of control messages flowed over the network
$\boldsymbol{k_i}$	Size of the control messages
$\boldsymbol{E_{Traffic}}$	Traffic load of a CH
$\boldsymbol{E_{Average\_load}}$	Average load between CH and its CM nodes
$\boldsymbol{R_i}$	The time period for one data gathering round
$\boldsymbol{n_g}$	Number of data packets receive by a node *j*
$\boldsymbol{E^j_{average}}$	The average energy consumption of a node *j*
$\boldsymbol{E^j_T}$	Energy consumed by a node *j* for data transmission
$\boldsymbol{E^j_{sen}}$	Energy consumed by a node *j* for data sensing
$\boldsymbol{B_x}$	The cluster region
$\boldsymbol{\vartheta}$	Width of a cluster region
$\boldsymbol{d_I}$	The distance between CH *i* and the BS
$\boldsymbol{E^i_{average\_CH}}$	The average energy consumption by CH *i*
$\boldsymbol{E^i_R}$	Energy consumed by CH *i* for data receiving
$\boldsymbol{E^i_T}$	Energy consumed by CH *i* for data transmitting
$\boldsymbol{E^i_{ED}}$	Energy consumption for data aggregation by i^{th} CH
$\boldsymbol{S_g}$	The set of strong nodes
$\boldsymbol{M_d}$	The set of moderate node
$\boldsymbol{W_k}$	The set of weak nodes
$\boldsymbol{E_{current}}$	Residual energy of a node

$\boldsymbol{E_{max}}$	Initial Energy of a node
$\mathbf{L_{avg}}(\boldsymbol{v_j})$	The average communication load of a CH *j*
$\boldsymbol{AC_{Load}}$	The average load of a cluster
$\boldsymbol{\delta_{avg}(v_R)}$	The average residual energy of the routing nodes
$\boldsymbol{TE_{max}}$	The total initial energy of the network

Table 2
Rule base used by the inference engine

Rule	***Communication_Load***	***Traffic_Load***	***Residual_Energy***	***Decision***
1	Low	Low	Low	Poor
2	Low	Low	Medium	Strong
3	Low	Low	High	Strong
4	Low	Medium	Low	Weak
5	Low	Medium	Medium	Moderate
6	Low	Medium	High	Moderate
7	Low	High	Low	Poor
8	Low	High	Medium	Weak
9	Low	High	High	Moderate
10	Medium	Low	Low	Weak
11	Medium	Low	Medium	Moderate
12	Medium	Low	High	Moderate
13	Medium	Medium	Low	Weak
14	Medium	Medium	Medium	Weak
15	Medium	Medium	High	Moderate
16	Medium	High	Low	Poor
17	Medium	High	Medium	Weak
18	Medium	High	High	Moderate
19	High	Low	Low	Poor
20	High	Low	Medium	Weak
21	High	Low	High	Weak
22	High	Medium	Low	Poor
23	High	Medium	Medium	Weak
24	High	Medium	High	Moderate
25	High	High	Low	Poor
26	High	High	Medium	Weak
27	High	High	High	Moderate

Table 3
Simulation parameters

Parameters	**Value**
Network size	100×100 [m^2]
Number of nodes	100-500 and 500-1000
τ_{fs}	10pJ/bit/m^2
τ_{amp}	0.0013 pJ/bit/m^4
τ_t	50 nJ/bit
$E_{agg}(P_i)$	5 nJ/bit/signal
Data packet size	100 bytes
Control message size	100 bits
Round time	30s
Frame in each round	5 frame
Threshold distance (d_0)	76 m
Initial energy	0.5 J
Radio propagation model	Log-distance path loss
Carrier sense range	150m
Duration of control period	5ms
Duration of data slot	12ms
Duration of RTS/CTS packet	0.9 ms
Duration of data packet	8.5ms
Channel rate	250kbps
Source rate	5pkt/s

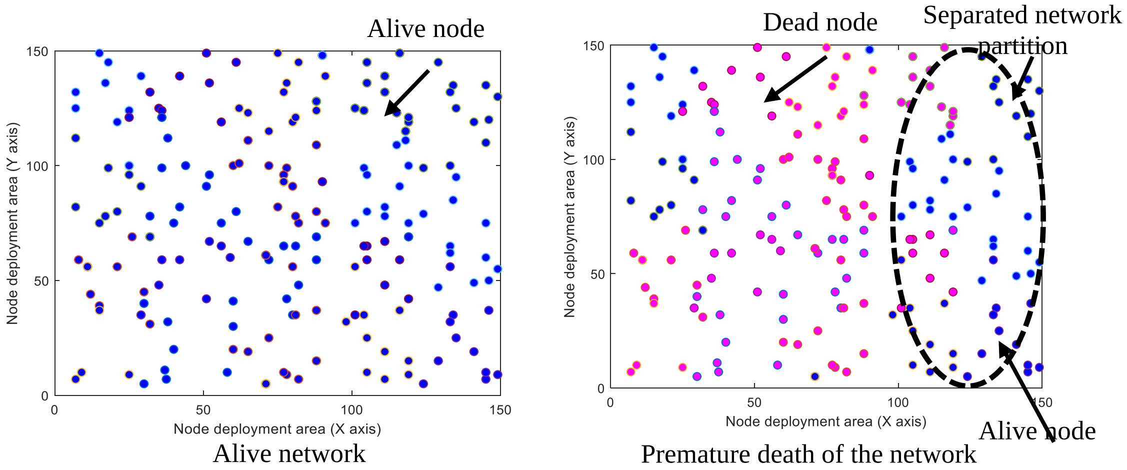

Fig. 1: Premature death of the dense WSNs.

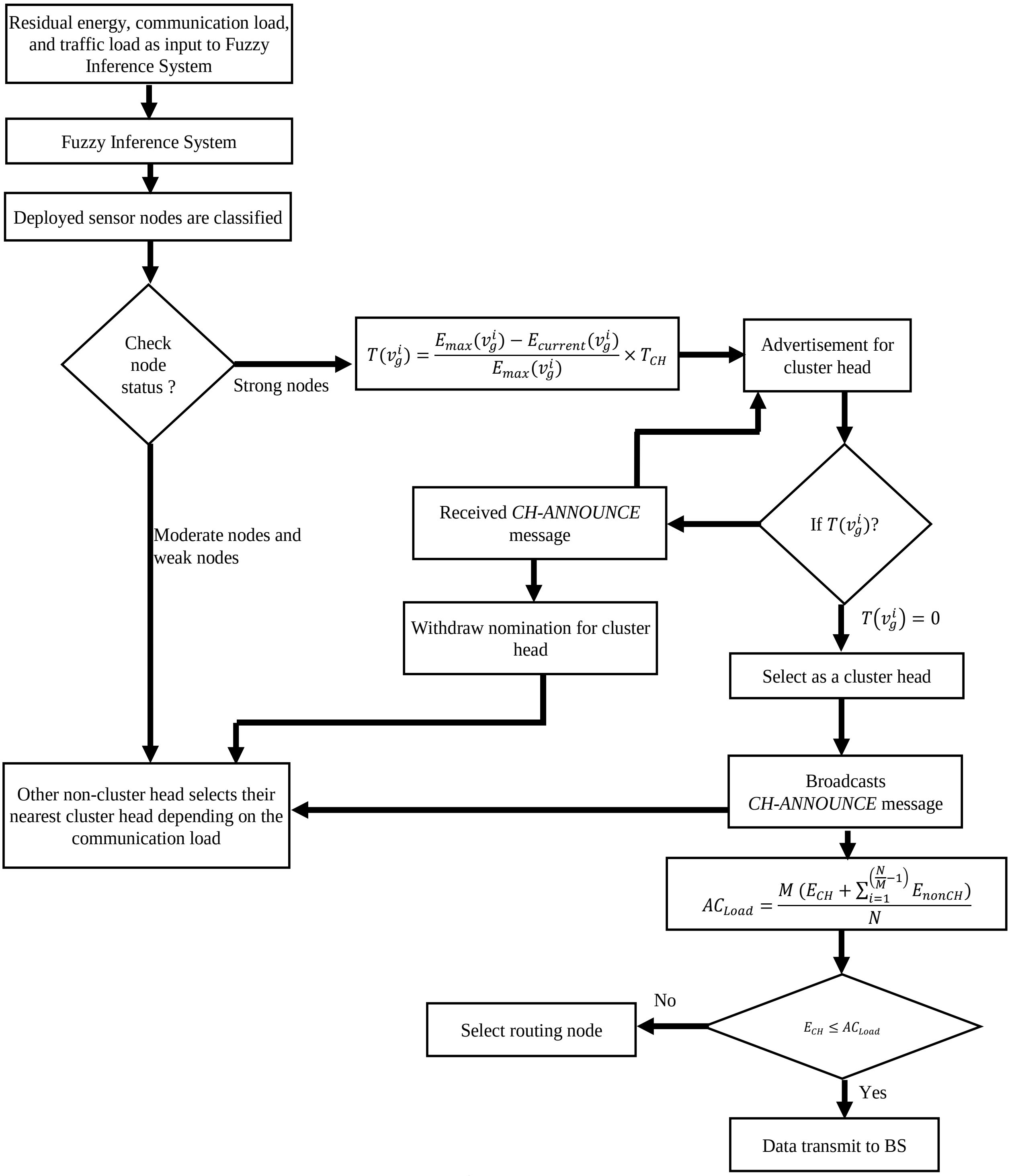

Fig. 2: Flowchart for the proposed scheme

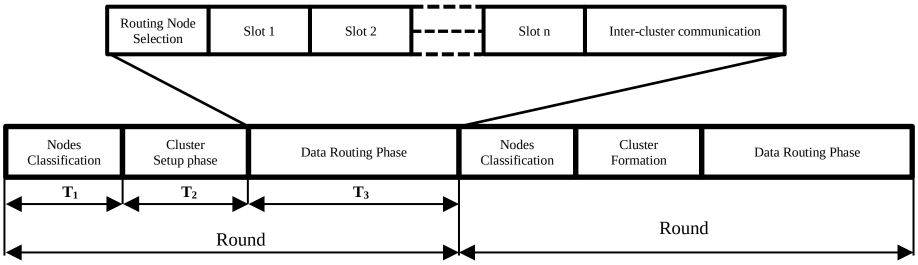

Fig. 3: Operation of the proposed scheme

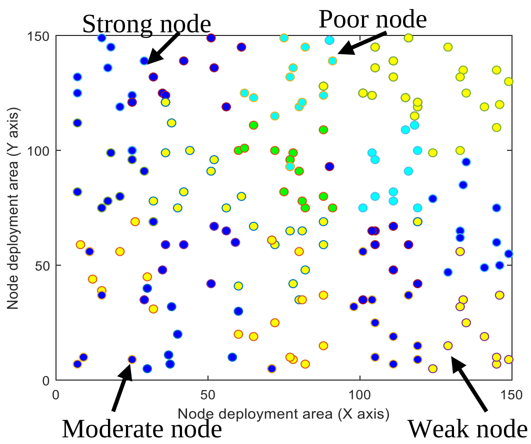

Fig. 4: Classification of nodes without dead node

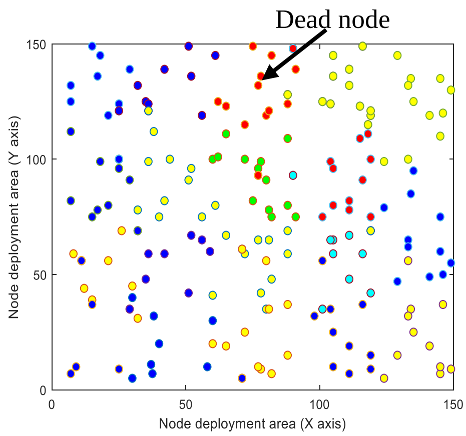

Fig. 5: Classification of nodes with dead node

FA	SID	RE	XPS	YPS

FA: Frame Identification
SID: Source ID
RE: Residual Energy
XPS: X Position of Source
YPS: Y Position of Source

Fig. 6: CH announcement packet (*CH_ANNOUNCE*) format.

FA	SID	RE	DBS	XPS	YPS

FA: Frame Identification
SID: Source ID
RE: Residual Energy
DBS: Distance from the BS
XPS: X Position of Source
YPS: Y Position of Source

Fig. 7: The routing node selection message (*TR_REQ*) format.

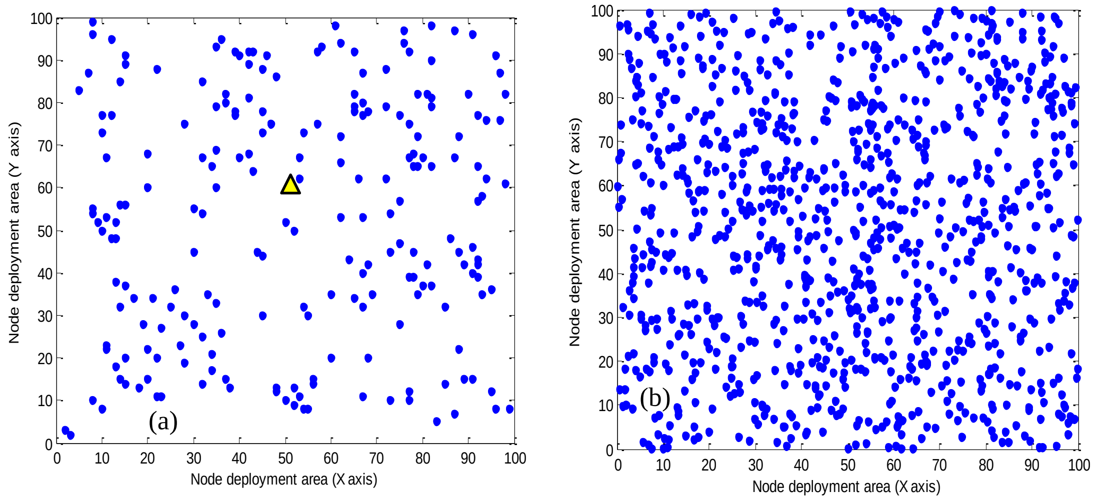

Fig. 8: Simulation scenarios. (a) Scenario 1 (low density network), and (b) Scenario 2 (dense network).

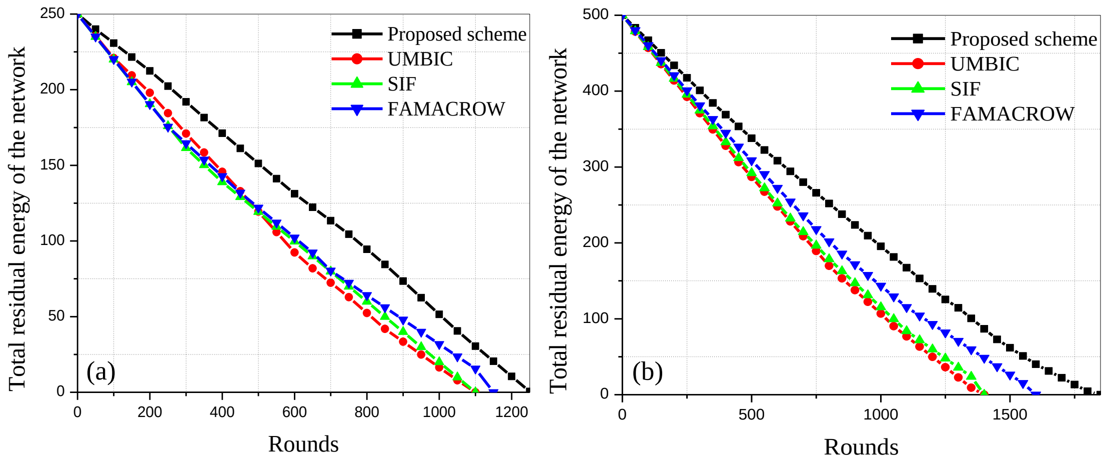

Fig. 9: Residual energy of the network in (a) Scenario 1 and (b) Scenario 2.

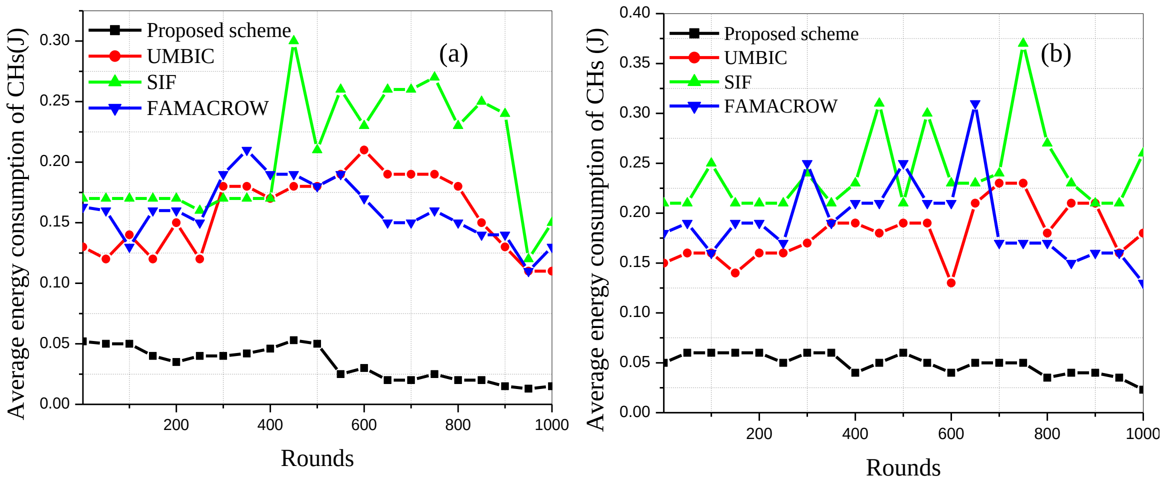

Fig. 10: Energy consumption of CHs in (a) Scenario 1 and (b) Scenario 2.

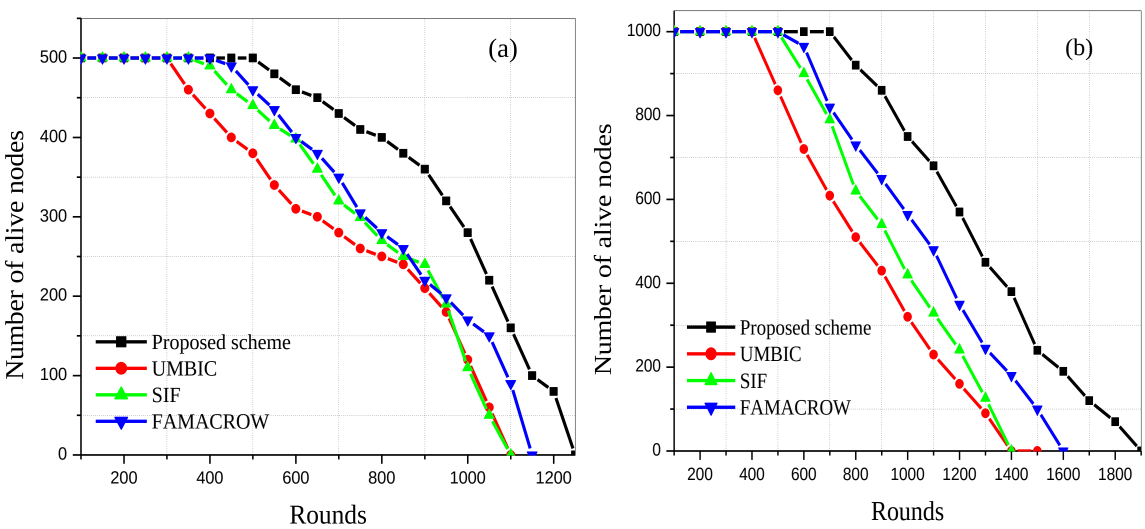

Fig. 11: Number of alive nodes over rounds, a) Scenario 1, b) Scenario 2.

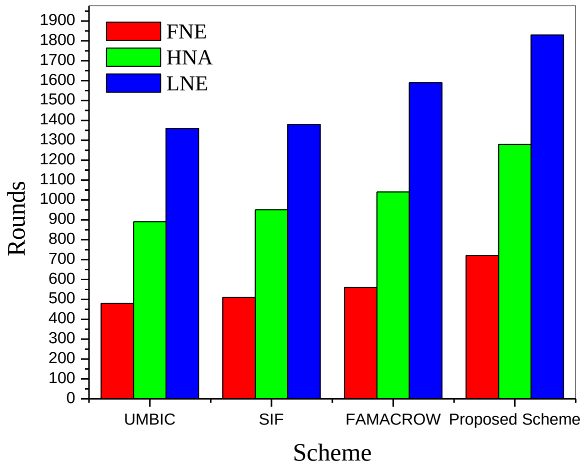

Fig. 12: (a) First Node Exhausted (FNE) in Scenario 1, (b) Half of nodes alive (HNA) in Scenario 1, (c) Last Node Exhausted (LNE) in Scenario 1.

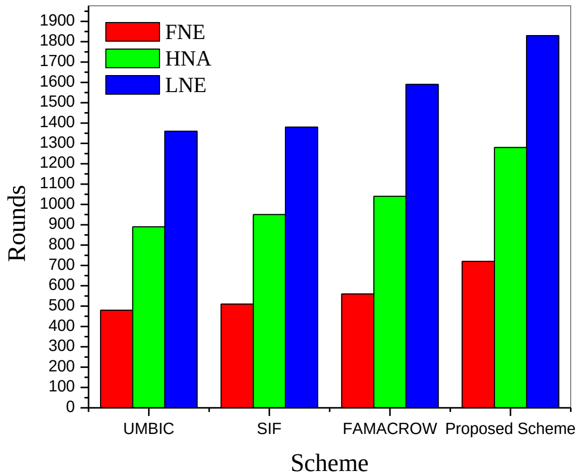

Fig. 13: (a) First Node Exhausted (FNE) in Scenario 2, (b) Half of nodes alive (HNA) in Scenario 2, (c) Last Node Exhausted (LNE) in Scenario 2.

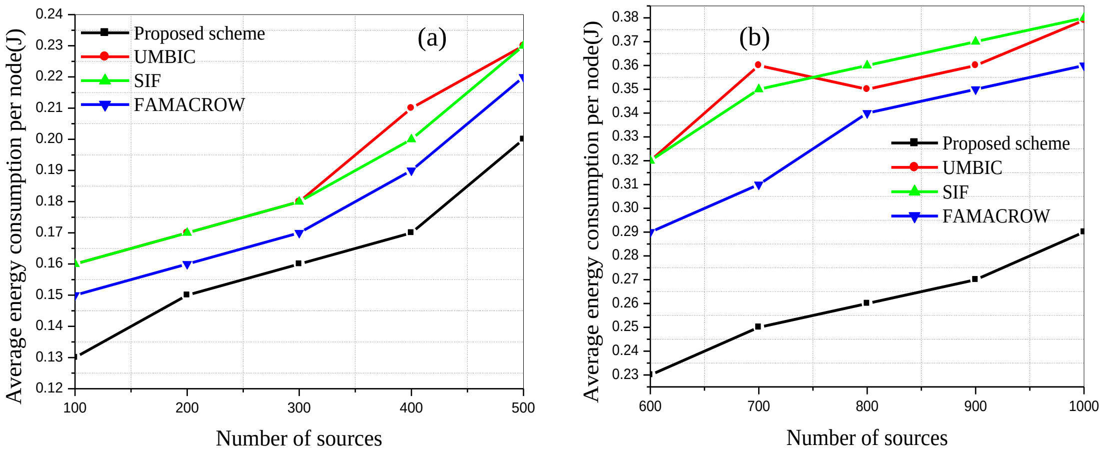

Fig. 14: Average energy consumption of the network, a) scenario 1, and b) scenario 2.

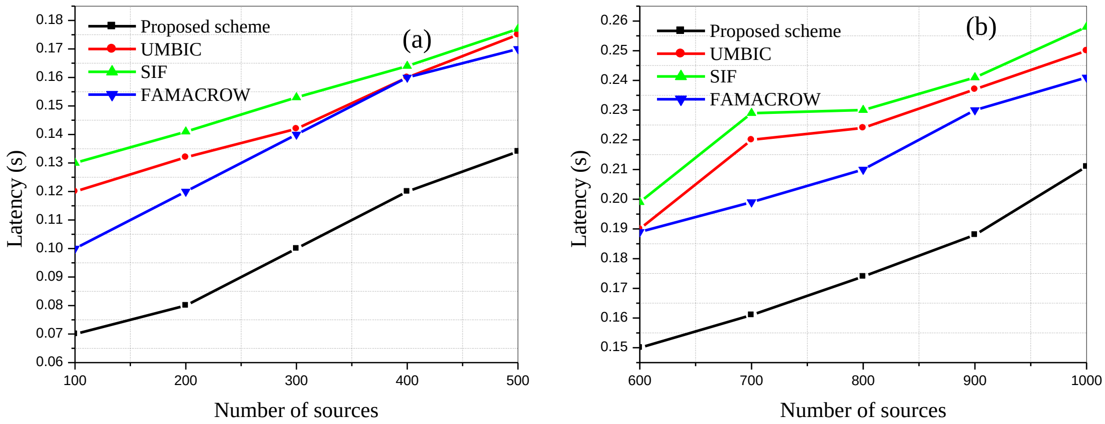

Fig. 15: Average data latency with different number of sources, a) scenario 1, and b) scenario 2.

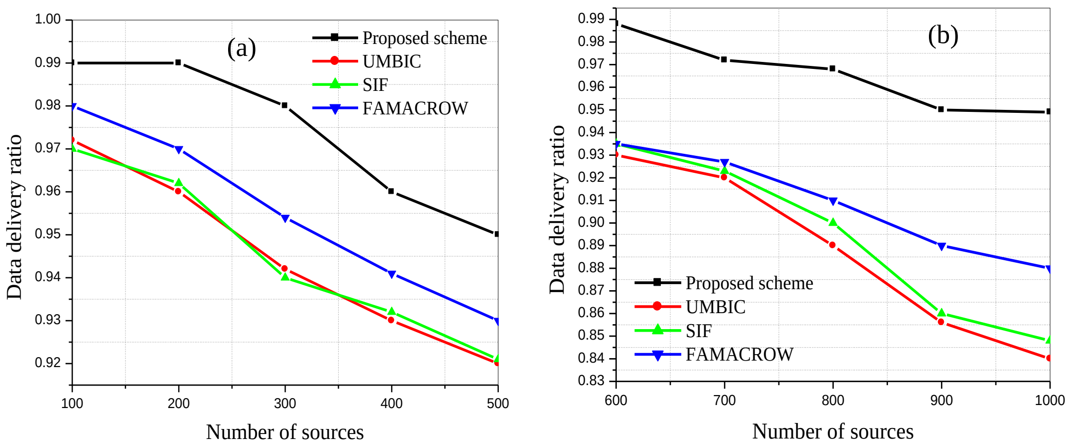

Fig. 16: Data delivery ratio with different number of sources, a) scenario 1, and b) scenario 2.