Data Redundancy-Control Energy-Efficient Multi-Hop Framework for Wireless Sensor Networks

Gulnaz Ahmed[1] · Xi Zhao[1] · Mian Muhammad Sadiq Fareed[2] ·
Muhammad Rizwan Asif[2] · Syed Ali Raza[3]

Abstract

Wireless Sensor Network (WSN) is an emerging technology that has attractive intelligent sensor-based applications. In these intelligent sensor-based networks, control-overhead management and elimination of redundant inner-network transmissions are still challenging because the current WSN protocols are not data redundancy-aware. The clustering architecture is an excellent choice for such challenges because it organizes control traffic, improves scalability, and reduces the network energy by reducing inner-network communication. However, the current clustering protocols periodically forward the data and consume more energy due to data redundancy. In this paper, we design a novel cluster-based redundant transmission control clustering framework that checks the redundancy of the data through the statistical tests with an appropriate degree of confidence. After that, the cluster-head separates and deletes the redundant data from the available data sets before sending it to the next level. We also designed a spatiotemporal multi-cast dynamic cluster-head role rotation that is capable of easily adjusting the non-associated cluster member nodes. Moreover, the designed framework carefully selects the forwarders based on the transmission strength and effectively eliminates the back-transmission problem. The proposed framework is compared with the recent schemes using different quality measures and we found that our proposed framework performs favorably against the existing schemes for all of the evaluation metrics.

Keywords Wireless sensor network · Data redundant · Control-overhead management · Cluster-based architecture · Best forwarder selection

✉ Xi Zhao
  zhaoxi1@mail.xjtu.edu.cn

✉ Mian Muhammad Sadiq Fareed
  sadiqfareed@mail.xjtu.edu.cn

1   School of Management, Xi'an Jiaotong University, Xi'an, China

2   School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China

3   School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China

1 Introduction

Wireless Sensor Network (WSN) is emerging as a potential technology that has attractive and updated applications. The intelligent smart-sensors utilized in WSNs are generally battery-operated. The smart-sensors are deployed in the sensing area to monitor the physical phenomena. After installation, these smart-sensors start working till the end of lifetime and the batteries of these smart-sensors cannot be replaced [1, 2]. Consequently, the WSNs sensors always face an unbalance energy issue. The network lifespan can only be increased by cutting down the energy consumption. If the energy consumption of these nodes is reduced, then these smart-sensors can stay alive for a longer time. However, this energy goal can be achieved by designing a protocol, which avoids long communication and data redundancy to minimize network energy consumption.

Different approaches [2] are designed to save the available resources of these smart-sensors. Among these approaches, the cluster-based architecture is more efficient, which saves the energy of the network by reducing the number of transmissions. Instead of forwarding the data individually, a head node is chosen for forwarding the sensed information of all the member nodes [3]. This sensed data is forwarded through a direct communication link or a cooperative communication link from the member nodes toward the surface sink. The cooperative communication links are preferred over the direct communication links when the data is forwarded to a long distance. The role of the head node revolves among all the member nodes. The head node performs some extra duties like data collection and data fusion [4, 5]. In performing such extra duties, the nodes consume some extra energy as compared to the other member nodes [2, 3]. However, the poor head node selection criteria, oversize clusters, and redundant data due to overpopulated clusters can deplete the battery of head node much earlier than expected.

In recent literature [6–9], different schemes and methods are proposed to avoid the similarity and redundancy in collected data. All these methods adopt different means to eliminate the similarity and to reduce the energy usage. To overcome the problem of energy consumption due to flooding-based routing, a method is defined in [6], where each node individually checks and eliminates the data packets which are not moving towards the destinations. A prefix frequency-filtering scheme is proposed in [7] to check the similarity between the data generated at the node level. This prefix frequency-filtering technique is also adopted in [8] to verify the similarity of the data sensed by neighbor nodes. An Euclidean and Cosine distance function-based scheme [9] is adopted to reduce the packet size and data redundancy. However, the redundancy is checked between the pair of sets that introduces data latency.

To overcome the issues of previously designed schemes, we introduce a novel Data Redundancy-controlled Energy Efficient Multi-hop (DREEM) clustering framework which increases the network lifetime by reducing the redundant and repeated transmissions over the link while ensuring the 99.9% data delivery at the surface sink. The CH selection is spatiotemporal multi-cast dynamic and capable of adjusting the non-associated cluster nodes due to inefficient clustering. The data-aggregation hierarchy is simple and data-forwarding routes are optimal, which also helps in improving the network lifetime. In our proposed framework, there is no need to employ all the nodes to take part in the sensing and collecting activities. Only the selected nodes from the different directions are chosen to perform the sensing. Consequently, the proposed framework proves to be very energy-efficient with less end-to-end delay and high delivery ratio as compared to the existing schemes. The main contributions of our proposed framework are summarized as:

- We propose a spatiotemporal multi-cast cluster-head selection that is based upon information delivery to the member nodes for a specified time limit in a defined region. Moreover, a dynamic cluster-head role rotation is defined that is capable of easily adjusting the non-associated cluster nodes due to inefficient clustering. In addition, these newly adjusted nodes can successfully forward their information to the CH according to their previous schedule.
- We propose a redundant transmission control framework that aims to eliminate the inner-network communication and the data-load at the base station. In this framework, every cluster-head applies a data redundancy check before forwarding data to the next level. This data redundancy check is based upon some statistical tests, these tests are performed with an appropriate degree of confidence and return a logical value after comparing the variance between their measures. If the tests results are positive, then the cluster-head separates and deletes the redundant data from the available data sets.
- We propose a clear strategy to avoid the back-transmission that increases the original path length and affect the network lifetime. In this strategy, the designed framework firstly determines a set of forwarders based on transmission strength and then, the Pythagoras theorem based distance calculation approach is adopted to decide the best forwarder towards the surface sink.
- We also investigate the network energy consumption of our framework to reveal the trade-off between energy consumption and data redundancy. From the simulation results, we found that the sensor nodes life is completely relying on the number of transmissions, by avoiding the redundant transmission and controlling the auxiliary information not to circulate inside the network can increase the $50 - 60\%$ lifetime of the network.

The remainder of this paper is organized in the following way. We discuss the current literature about the WSN in Sect. 2. In Sect. 3, each stage of the proposed model like network initialization, cluster head selection, and data collection are discussed in detail. In Sect. 4, the simulation analysis and comparison with state-of-the-art is described to examine our proposed model. In Sect. 5, the conclusion of the proposed framework is drawn.

2 Related Work

To equally divide the energy load and balanced the energy consumption a layers based clustering architecture is defined in [10]. The clusters size increase gradually as the distance increases from the BS. The cluster radius, the connectivity parameter, and the nodes density are the basic parameters used for dividing the networks into the clusters. The main purpose of this framework is to produce the different size clusters, which is free from excessive node responsibilities and processing. In order to alleviate the burden of CHs, a centralized algorithm is discussed in [11]. The BS is a central unit for selecting the CHs, because the BS has more sophisticated hardware as compared to the ordinary nodes. The CH selection process is based on the behavior of the nodes in the previous round. Artificial bee colony algorithm is used to find the optimal solution for data collection and data forwarding.

To overcome the current routing problems, a routing protocol based on the Tabu-search optimization is defined in [12]. A routing-cost related function is computed based on residual energy, transmission cost, and the number of hops towards the BS. This

method updates the Tabu-list and the Tabu-tenure in each round by picking up the high energy nodes. This Tabu search method utilizes the next-hop neighbor information and neighbor distance from BS as a substitute in case of route failure. A multi-layer heterogeneous protocol is designed in [13] to increase the network stability on the belief that some node consumes more energy at some particular places. They mixed a percentage of higher energy nodes with lower energy to create a sensor network. They check the performance of the network with different percentages of the energy heterogeneity and found that the network stability is dependent on the heterogeneity factor. However, they neglected that by mixing a higher percentage of energy heterogeneous nodes increase the overall network energy as compared to the opponents.

In [14], a novel method is designed for partitioning the network into small equal size fan-shaped clusters. To acquire fan-shaped clustering, the network sensing field is first divided into concentric circles. After that, each circle is further partitioned into equal parts to obtain the fan-shaped fixed clusters. Once a node is elected as a CH, it remains a CH until it's energy-level reaches to a certain threshold. However, the BS closer CH forwarding the information of more than nine-clusters contains more than 700 sensor nodes. These CHs deplete their batteries more quicker than other distant CHs. A virtual ring-based routing method is engaged in [15], this method considers two types of nodes like powerful nodes and weak nodes. The powerful nodes work as CH and use the weak nodes to sense the network sensing field after a specific time slot. These clever CH nodes work as head until they have predefined high energy value, after that these nodes also working as the member nodes.

The uneven clustering is more cooperative in comparison with even and flat clustering in saving the network resources [16]. The authors divide the closer regions into small clusters and far regions into big clusters depending on the distance from the main station. In [17], a multi-level clustering based algorithm is defined for WSNs. In this algorithm, the nodes with higher energy are chosen as CHs, while the lower energy nodes are act as the cluster member nodes. The CHs use the communication radius to approach the lower energy nodes and assign them task of sensing. The lower energy nodes task end after sending data to high energy nodes (CHs) while the CHs finish their monitoring task after handing over the data to the BS. In [18], a Hybrid Energy-Efficient Multi-path routing Protocol (HEEMP) is designed to increase the performance and quality of services of clustering architecture. The CHs are chosen by the BS using the node degree and remaining energy of the nodes. They restricted the cluster members to communicate within the clusters in a multi-hop manner either the cluster size is small or large. However, due to this restriction the back-off communication starts in the clusters and a packet travel a long distance to reach the CH as compared to normal path. A Distributed Fault-tolerant Clustering and Routing (DFCR) is designed in [19] to overcome the clustering and routing faults in WSN. The nodes that are not the part of clustering architecture either as a CH or a member node due to some unseen factors, or adjusted in the system using distributed runtime recovery algorithm. The CHs convey the data using the multi-hop communication and always save neighbor information as a backup in the case of route failure.

The proposed model is compared with recent cluster-based routing approaches on the basis of clustering objective, the ideal clustering properties, and clustering capabilities in Table 1. From this comparison, we found that the proposed model have excellent characteristics and best clustering objectives as compared with current clustering schemes. The limitations of the proposed models are given as:

Table 1 Clustering properties and cluster capabilities comparison of the recent clustering protocols

Method	Cluster over-lapping	Time of coverage	Dealing with FCHs	Location awareness	Optimal	Type of node	Energy-efficient	Route recovery	CHs selection
FSC	No	Fixed	No	Yes	No	Homogeneous	Yes	No	Fixed
LEMHR	Yes	Fixed	No	No	No	Heterogeneous	No	No	Fixed
ENEFC	No	Variable	No	No	No	Homogeneous	Yes	No	Fixed
ABC	No	Variable	No	No	No	Homogeneous	Yes	Yes	Random
MHCOA	OK	Variable	No	No	No	Heterogeneous	Yes	No	Random
CLEVER	No	Fixed	No	No	No	Homogeneous	Yes	No	Random
SEED	OK	Variable	Yes	Yes	No	Homogeneous	Yes	Yes	Fixed
HEEMP	No	Variable	No	No	No	Homogeneous	Yes	No	Random
DREEM	No	Variable	Yes	Yes	Yes	Heterogeneous	Yes	Yes	Fixed

- The Big Cluster-heads (BCHs) are only relaying (working as a forwarder) the data of Small Cluster-heads (SCHs) in their Big Cube (BC) and not taking part in any sensing activity.
- As the proposed model is redundancy-aware, sometimes the normal data in overlapping region is considered as redundant and deleted.
- The proposed model support the large-scale WSNs, however, the proposed model does not perform persuasively for very large-scale WSNs. Because for the very large-scale networks the CHs remains busy in forwarding the data of their predecessor. When we significantly increase the number of nodes in the network this significantly increase the data-load on these CHs. These CH nodes have limited capabilities, and then they convey a limited data to the BS according to their capabilities and remaining is lost or these CHs deplete their batteries earlier making the network unstable.

3 Energy Efficient Redundant Transmission Control Framework

To explain the operation of our framework, we divide it's working into rounds (time steps). Then, every round of DREEM is further divided into four steps such as; (1) network initialization, (2) Small Cluster-Head (selection, (3) data collection at SCH and redundancy elimination, and (4) data collection at Big Cluster-Head and best forwarder selection as shown in Fig. 1.

3.1 Network Initialization

In our proposed framework, the network-sensing field is divided into Big Clusters (BCs) and each BC is further divided into Small Clusters (SCs) as discussed previously. The edge length of a BC is considered as R and the edge length of a SC is considered as r. The length of the SC is adjusted according to the communication range of nodes. Each node is represented through it's location $i(a, b, c)$ and SC number $N(o, p, q)$. While the o, p, and q can be calculated by the following equations:

$$o = \frac{r - a \quad \mod r + a}{r} \tag{1}$$

$$p = \frac{r - b \quad \mod r + b}{r} \tag{2}$$

$$q = \frac{r - |c| \quad \mod r + |c|}{r} \tag{3}$$

After the completion of network configuration, BCH broadcasts the length of SC's edge r and the coordinates of BCH to all the nodes. According to the Eq. (21), each node can calculate it's distance $d_{i,BCH}$ from the other node and BCH by using the following equation:

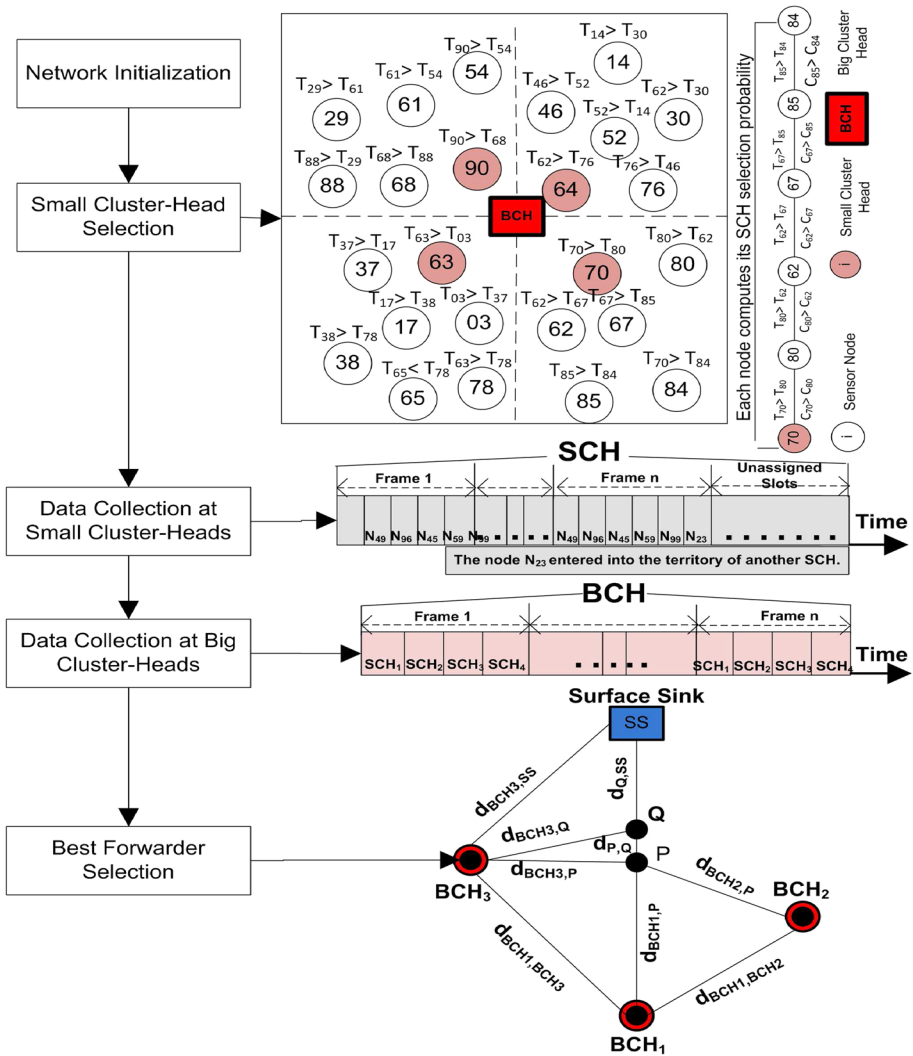$$d_{i,j} = \sqrt{(a_i - a_j)^2 + (b_i - b_j)^2 + (c_i - c_j)^2} \tag{4}$$

Fig. 1 The flowchart of our proposed framework containing the details about the data collection at SCHs and the best forwarder selection

Before discussing the operations of our proposed framework, we are making some assumptions as follows:

- The BS is an enriched resources device. It is equipped with the processor with high processing speed and unlimited energy resources.
- The BCHs are only relaying (working as a forwarder) the data of SCHs in their BC and not taking part in any sensing activity.
- All the sensor nodes including the SCHs contain the equal amount of initial energies.

- The communication channel used is symmetric. It's mean that the power consumed by N_i to forward information to N_j is equal to the energy consumed by N_j to convey it's data to N_i for a defined signal to noise ratio.
- The nodes in the overlapping region contain the correlated data.
- Nodes can perform the sensing duties in sleep-mode or low-power listening mode, but they can communicate only in the active-mode.

3.2 Small Cluster-Heads Selection

The SCH selection is the main phase in which a head node is chosen that is responsible for collecting the data from the member nodes and successfully forwarding it to the next level. The SCH selection phase starts after the network initialization when each BCH sends a control packet to all the nodes in it's associated BC. The details of SCH selection phase are given as follows:

- The node will start a timer t to 1 when it deployed to the sensing field and also when the SCH selection is started. This timer is calculated as:

$$t_i = t_{max} \times \frac{R_{EN}}{I_{EN}} \times \frac{d_{F,BCH} - d_{i,BCH}}{d_{F,BCH}} \tag{5}$$

$$d_{F,BCH} = r\sqrt{(o^2 + p^2 + q^2)} \tag{6}$$

where R_{EN} and I_{EN} are the residual and the initial energies of the node, respectively. $d_{i,BCH}$ is the distance of $NodeN_i$ from the BCH and $d_{F,BCH}$ is the maximum distance of the SC from the BCH. In the beginning, all the nodes have the same amount of initial energies. So, the nodes closer to BCH likely have more chance to be selected as SCH but with time, all the members within the same cube have equal chances to be selected as SCHs.

- After that, each node broadcasts a SCH selection message *ADV_SCH_SELECT* at the range $\sqrt{3}r$. This message *ADV_SCH_SELECT* contains: node's SC identifier (o, p, q), node coordinates (a, b, c), node's distance from the BCH, the initial and the residual energies of the node.
- The node that receives this SCH selection message *ADV_SCH_SELECT* checks either this message belongs to same SC or not. if it belongs to the same SC, then it further analyses this message and checks the possibility of this node as a SCH through this equation:

$$SCH_i = \frac{R_{En}}{d_{i,BCH}} \tag{7}$$

The node then computes and compares the SCH selection probability of that node SCH_j with its own SCH_i. If the SCH_j of the received message is lesser, then the SCH_i this node information is saved to the cluster table. An example is described in Table 2 in which a SCH N_{70} maintained its cluster table.

- If any of the nodes receives a message *ADV_SCH_SELECT* after the time is ended and the SCH is not selected yet.

Table 2 Cluster table maintained by SCH N_{70}

Node ID i	Node position (a, b, c)	SC ID (o, p, q)	Node residual energy R_{EN}	Node distance to BCH $d_{i,BCH}$	SCH selection measure SCH_i
70	(8, 15, 16)	(4, 4, 4)	2000	17	117.64
80	(8, 16, 15)	(4, 4, 4)	2000	18.5	108.10
62	(9, 15, 13)	(4, 4, 4)	2000	19.5	102.56
67	(9, 14, 12)	(4, 4, 4)	2000	22	90.90
84	(10, 12, 14)	(4, 4, 4)	2000	24	83.33
85	(10, 12, 17)	(4, 4, 4)	2000	25	80.00

After the SCH selection process, each SCH sends a control packet to all the nodes in its SC. The field nodes that receive this control packet send a joint request to the SCH. On receiving the joint request, the SCH assigns each member node with a unique ID for future correspondence. After receiving the node IDs, the member nodes switch to sleep-mode and wait for a time division multiple-access slot.

3.3 Data Collection at Small Cluster-Heads

After the successful completion of SCH selection phase, the TDMA slot allocation starts. The SCH selects nodes from the different direction and assigns them the TDMA slots. For this selection, only those nodes are chosen that are not either SCHs nor perform the sensing activity in previous rounds. The nodes that do not perform the sensing activity remain in sleep-mode to save the available resources. While the chosen nodes start sensing their fields, at the arrival of their time slots these field nodes convey their sensed information to SCHs.

3.3.1 Variance Study

In the beginning, we execute a number of statistical tests to analyze the means of these tests are equal or not. In these tests, we suppose that variance in these data sets is not very significant. Therefore, the S_{out} is computed in a different way through the statistical tests and the S_{out} is computed as a ratio in the variance depending upon the measurements. For every false rejection probability, the sets are reproduced if the S_{out} is less than the threshold T_{DOF}.

3.3.2 Assumption and Definitions for Variance Test

- Suppose $N = \{N_1, N_2, ..., N_n\}$ represent a set of member nodes that are generating a data set $S = \{S_1, S_2, ..., S_n\}$ in a specified time slot.
- Suppose $SCHs = \{SCH_1, SCH_2, ..., SCH_l\}$ represent the set of SCHs, where $l \leq n$. Every time the SCH collects n data sets in its SC from the member nodes.
- Each time the received data set contains T number of measures.
- We also suppose that the number of measures in every data set $|S_j|$ is independent from the mean $\overline{X_i}$ while $\sigma_n^2 = \sigma$.

Definition 1 The similar function is defined as when a node captures two functions with similar measures and express as follows:

$$Similar(s_i, s_j) = \begin{cases} 1 & if \ \|s_i, s_j\| \leq \delta \\ 0 & Otherwise \end{cases} \tag{8}$$

where $s_i, s_j \in S$ and δ is the threshold value.

Definition 2 The weight for a measurement s_i can be defined as: it is the occurrence of a similar function in the alike set.

Definition 3 The cardinality of a set can be defined as: the cardinality of a set S_n is equivalent to its number of rudiments.

Definition 4 Weighted cardinality can be defined as: the weighted cardinality of set S_n is equivalent to the sum of all measure's weight in the set $W_{card}(S_n)$. Now the measure's variable can be written [20, 21] as:

$$s_{ji} = \overline{X_i} + \epsilon_{ji}; \qquad j = 1, ..., n \quad and \quad i = 1, ..., |S_i| \tag{9}$$

where, ϵ_{ji} is the residual and independent and follow the normal distribution $N(0, \sigma^2)$. For each data set $|S_j|$, we represent $\overline{X_i}$ as the mean, σ_j^2 as variance, and $\overline{X}$ as the mean of available data sets, respectively.

$$\overline{X_j} = \frac{1}{W_{card}(S_j)} \sum_{k=1}^{|S_j|} \left(s_{jk} \times W(s_{jk}) \right) \tag{10}$$

$$\sigma_j^2 = \frac{1}{W_{card}(S_j)} \sum_{k=1}^{|S_j|} \left(\left(s_{jk} - \overline{X_j} \right)^2 \times W(s_{jk}) \right) \tag{11}$$

$$\overline{X} = \sum_{j=1}^{n} \sum_{k=1}^{|S_j|} \frac{\left(s_{jk} \times W(s_{jk}) \right)}{W_{card}(s_j)} \tag{12}$$

where, $s_{jk} \in S_j$ and $W(s_{jk}$ is measur's weight. Since, $W_{card}(S_1) = ... = W_{card}(S_i) = ... = W_{card}(S_n) = T$

$$\overline{X_j} = \frac{1}{T} \sum_{k=1}^{|S_j|} \left(s_{jk} \times W(s_{jk}) \right) \tag{13}$$

$$\sigma_i^2 = \frac{1}{T} \sum_{k=1}^{|S_j|} \left(\left(s_{jk} - \overline{X_j} \right)^2 \times W(s_{jk}) \right) \tag{14}$$

$$\overline{X} = \frac{1}{T} \sum_{j=1}^{n} \sum_{k=1}^{|S_j|} \left(s_{jk} \times W(s_{jk}) \right) \tag{15}$$

We computed the overall variation within a set and in between the sets. In the beginning, we calculated the mean of the given measurement and after that variances of all other sets is compared with these means. We use our previous hypothesis that the variance in the sets is not significant. So, the weighted variance among the sets will be considered same as variance in the sets.

3.3.3 Honestly Significant Difference Test

In this part, we perform the honestly significant difference test [22] to calculate the means and variances for the data sets. After this test, we decide either these data sets are redundant or not.

$$T_{SOS} = \sum_{j=1}^{n} \sum_{k=1}^{|S_j|} \left(s_{jk} \times W(s_{jk}) \right)^2 - \frac{\left(\sum_{j=1}^{n} \sum_{k=1}^{|S_j|} s_{jk} \times W(s_{jk}) \right)^2}{n \times T} \tag{16}$$

$$SOS_{between} = \frac{\sum_{j=1}^{n} \left(\sum_{k=1}^{|S_j|} \left(s_{jk} \times W(s_{jk}) \right)^2 \right)}{T} - \frac{\left(\sum_{j=1}^{n} \sum_{k=1}^{|S_j|} s_{jk} \times W(s_{jk}) \right)^2}{n \times T} \tag{17}$$

$$SOS_{inside} = T_{SOS} - SOS_{between}; \quad DOF_{between} = n - 1; \quad DOF_{inside} = n(T - 1) \tag{18}$$

$$MOS_{between} = \frac{SOS_{between}}{DOF_{between}} \tag{19}$$

$$MOS_{inside} = \frac{SOS_{inside}}{DOF_{inside}} \tag{20}$$

$$S_{out} = \frac{MOS_{between}}{MOS_{inside}} \tag{21}$$

$$T_{DOF} = DOF(DOF_{between}, DOF_{inside}) \tag{22}$$

So, whenever we are conducting honestly significant difference test, we have to confirm that the value of S_{out} should be significant within the probability table with an appropriate degree of freedom $T_{DOF} = DOF(DOF_{between}, DOF_{inside})$. The result depends upon S_{out} and T_{DOF}:

- If the value of S_{out} is greater than T_{DOF}, then our assumption is discarded due to false rejection probability α, while the variance is considered significant among the data sets.
- if S_{out} is lesser or equal to the T_{DOF}, then our assumption is considered valid.

3.3.4 Redundancy Elimination at Small Cluster-Heads

Based on the previous variance tests, the data with low variance in their measure is selected. To check the redundancy of the data, the proposed Algorithm tests and returns a Boolean value by using the selected data. Initially, this Algorithm calculates the value of S_{out} and look for the threshold value T_{DOF}, which is calculated through the statistical tests with an appropriate degree of confidence. Lastly, it returns a logical value after comparing the threshold value is greater than the variance between their measures. After all these tests and confirmation, the SCH node separates and deletes the redundant data from all the available data. This algorithm also demonstrates how the SCH elects the data to be conveyed from all the available data sets, and which data should be forwarded to the BCH. Instead of all available data, only the data with the highest measure is sent to the BCH to improve the system efficiency and save the available resources.

3.4 Data Collection at Big Cluster-Heads

Each BC contains a BCH that is enriched with high capabilities as compared to the other sensor nodes. The BCHs are only relaying the data of SCHs and not taking part in any sensing activity. There is no need for time scheduling to forward the data to the BCH. SCHs can forward the data to their corresponding BCHs anytime during a round. The BCH collects the data from the SCHs in a BC and forwards it to the next BCH toward the surface sink. The forwarder BCH selection criteria are discussed in the next subsection.

3.4.1 Best Forwarder Selection

To solve the best forwarder selection, we take a scenario in which a BCH_1 has data to forward. The BCH_1 has two choices to forward the data toward the surface sink BCH_2 and BCH_3. The BCH_1 measures its distance from the surface sink $d_{BCH_1,SS}$, and BCH_1 calculates a middle point P from the surface sink. After that, the BCH_1 determines the signal strength of both the BCH_2 and BCH_3 at P. The BCH with higher signal strength at point P will be selected as a forwarder. Although the BCH_3 is at the longer distance from the BCH_1, but this BCH_3 avoids the back transmission and helps in saving the energy of the network. To explain this scenario, we develop this formulation:

$$(d_{BCH_3,SS})^2 = (d_{BCH_3,Q})^2 + \left(\frac{d_{BCH_1,SS}}{2} - d_{P,Q}\right)^2 \tag{23}$$

whereas $d_{BCH_3,Q}$ is the distance between BCH_3 and the point Q, whereas Q is a midpoint between the line joining point P to the BS, while $d_{P,Q}$ is the distance of point P from the point Q.

$$(d_{BCH_1,BCH_3})^2 = (d_{BCH_3,Q})^2 + \left(\frac{d_{BCH_1,SS}}{2} + d_{P,Q}\right)^2 \tag{24}$$

whereas $d_{(N,SS)} = d_{(N,MP)} + d_{(MP,SS)}$, by adding Eqs. (23) and (24). Then, we have the following expression:

$$(d_{BCH_3,SS})^2 + (d_{BCH_1,BCH_3})^2 = 2(d_{BCH_3,Q})^2 + \frac{(d_{BCH_1,SS})^2}{2} + 2(d_{P,Q})^2 \qquad (25)$$

After putting the values, $(d_{P,Q})^2 = (d_{BCH_2,P})^2 - (d_{BCH_3,Q})^2$

$$(d_{BCH_3,SS})^2 + (d_{BCH_1,BCH_3})^2 = \frac{(d_{BCH_1,SS})^2}{2} + 2(d_{P,BCH_3})^2 \qquad (26)$$

We can see from the above Equation, the distance between the BCH_1 and surface sink is fixed, while this part of the Equation $(d_{BCH_3,SS})^2 + (d_{BCH_1,BCH_3})^2$ is directly depending upon the $(d_{P,BCH_3})^2$. If we minimize this part of the Equation $(d_{BCH_3,SS})^2 + (d_{BCH_1,BCH_3})^2$, then we can minimize this part of the Equation $2(d_{P,BCH_3})^2$, and we can achieve our objective. As a result, if a BCH chooses a forwarder closer to the middle point in the direction of the surface sink. In that case, the communication distance becomes smaller this can help in saving the energy of the network.

4 Performance Analysis

We perform simulation to assess our proposed model with the current models MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18]. We use the different evaluation metrics to analyze the performance of DREEM. The selected approaches are most recent and functioning of these approaches is similar to DREEM. For a fair comparison, the same simulation environment and same parameters are used for all these approaches. Simulation parameters applied for the experiments are specified in Table 3.

Table 3 Parameter values used for the simulations

Variable	Value
Number of field nodes	100–500
Network area	$100\,m \times 100\,m$
r	10 m–50 m
Transmission range	250 m
Transmit power	50 W
Bandwidth	80 Hz
E_{int}	1000–3000 J
E_{elec}	50 nJ/bit
$A(f)$	1.001
d_0	80–100 m
E_{da}	50 nJ/bit/packet
Data rate	5 Kb/s
Data packet size	64 bytes
Header size	13 bytes
Total run time	10000 s

(a) Aggregated measure percentage with different values of α and T.

(b) The energy consumption comparison of our framework with and without data redundancy.
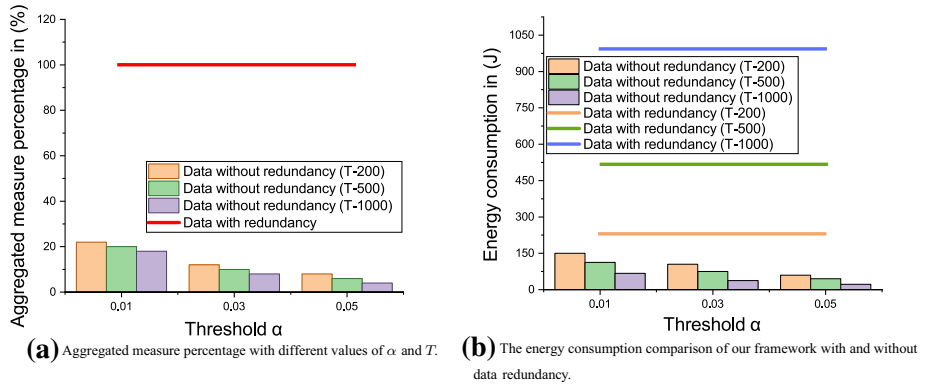
Fig. 2 The effect of data redundancy on the performance of our proposed framework with different values of α and T

4.1 Our Framework Performance with Changing α and T

The sensors are deployed in the network are application specific, so mostly the sensed information is similar. Figure 2a represents the amount of data transmitted towards the surface sink with and without redundancy. In the worst scenario $\alpha = 0.01$ and $T = 200$, where only 25% of collected data is forwarded to the surface sink. From this figure, we can note that by increasing the α and T, we reduce more amount of redundant data and only 10% of the collected data is traveled towards the surface sink.

Figure 2b shows the effect of redundancy on energy consumption with different values of T and α. From the results shown in this figure, we found that the energy consumption of the SCH or the field nodes is highly reliant on the amount of data. As the proposed framework is redundancy-aware, so it significantly reduces the amount of redundant collected data of the member nodes and only selected data is forwarded to the surface sink. Consequently, it saves a good proportion of available energy resources as revealed in Fig. 3b.
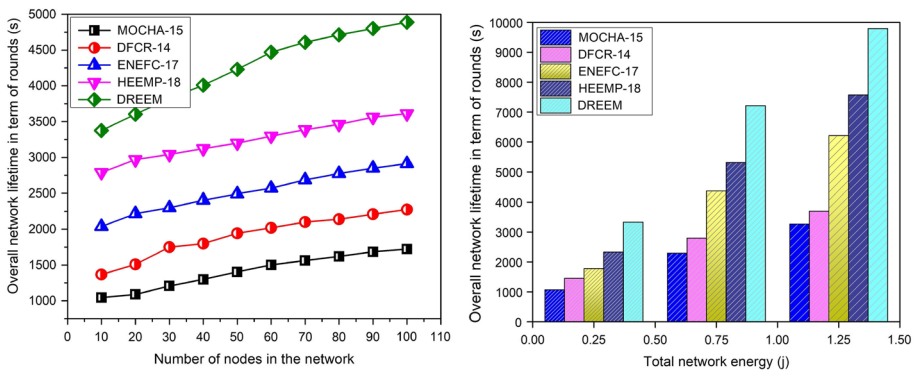

Fig. 3 The network energy consumption comparison of DREEM with current approaches using different initial energies

4.2 The Lifetime of the Network

Figure 3a demonstrates the overall network lifespan comparison of the DREEM against the recent clustering models MOCHA, DFCR, ENEFC, and HEEMP. From Fig. 3a, we can see that the lifespan of MOCHA, DFCR, ENEFC, HEEMP, and DREEM is 2179, 3042, 3695, 3855 and 4919 rounds (time steps), respectively. The overall lifespan of DREEM is 55%, 39%, 25%, and 22% greater than MOCHA, DFCR, ENEFC, and HEEMP, respectively. Network lifetime of DREEM is 2740, 1877, 1224, and 1064 rounds greater than MOCHA, DFCR, ENEFC, and HEEMP, respectively better because these models do not deal with data redundancy. According to our simulation results 50-60% of total data circulate inside the network is redundant or similar. These models just forward the received packet to the next hop without confirmation. However, some recent models erase the data packet with the same ID. This redundant data is the loss of network resources and decreases the network lifetime. We also perform the simulation with different initial energies and node distribution. Figure 3b depicts the performance of MOCHA, DFCR, ENEFC, HEEMP, and DREEM with dissimilar total network energies. Tables 4 and 5 show the lifetime comparison of MOCHA, DFCR, ENEFC, and HEEMP with node density $N = 300$ and $N = 500$. The DREEM remains consistent and performed favorably as compared to the MOCHA, DFCR, ENEFC, and HEEMP in all scenarios.

4.3 The Network Energy Consumption in Stable and Unstable States

The state of the network before its first field node drains its battery is network stable state, the network performs well in the stable state after that unstable network state starts till the end of the network. Figure 4a demonstrates the network energy comparison of the DREEM against the recent clustering models MOCHA, DFCR, ENEFC, and HEEMP. From the Fig. 3a, we can see that the stable state of DREEM is 2035, 1511, 1029 and 395 rounds greater than MOCHA, DFCR, ENEFC, and HEEMP, respectively. The stable state of DREEM is 64%, 58%, 23%, and 13% greater thanMOCHA, DFCR, ENEFC, and HEEMP, respectively because these models do not deal with data-redundancy. So, the redundant and similar data packets and control packets freely circulate across the network which leads to greater energy consumption as compared to DREEM. Figure 4b is network stability period comparison DREEM, MOCHA, DFCR, ENEFC, and HEEMP. The DREEM performs persuasively against the others with changing the initial energy-level. Tables 6 and 7 show the energy comparison with different node distributions $N = 300$ and $N = 500$. It is worth noting that the DREEM performs equally well in different node densities as compared to MOCHA, DFCR, ENEFC, and HEEMP.

4.4 The Network Throughput

Figure 5a shows the comparison between the network throughput and the network lifetime for the DREEM, MOCHA, DFCR, ENEFC, and HEEMP. We can note that the throughput of the proposed model is not persuasive as compared to the other models because the DREEM is redundancy-aware and only those data packets travel toward the BS that have not any similarity with previous data packets. DREEM checks the redundancy of the data through the statistical tests with an appropriate degree of confidence. Initially, our proposed model computes the values of the measures and returns a logical value after comparing the

Table 4 Lifetime comparison of DREEM with MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18] over 5 simulations (with 300 nodes)

WSN#	MOCHA-15			HEEMP-18			ENEFC-17			DREEM		
	FND	HND	LND	FND	HND	LND	FND	HND	LND	FND	HND	LND
1	1055	1592	2018	2805	3433	3808	2162	2802	3716	3196	4203	4784
2	1065	1551	2101	2818	3416	3797	2152	2777	3760	3188	4168	4829
3	1100	1494	2083	2855	3566	3845	2135	2796	3694	3159	4161	4794
4	1133	1484	1999	2809	3372	3822	2180	2816	3729	3111	4146	4866
5	1099	1572	1969	2785	3500	3811	2212	2840	3642	3207	4246	4907
Average	1090.4	1538.6	2034	2814.4	3457.4	3816.6	2168.2	2806.2	3708.2	3172.2	4184.8	4836

Table 5 Lifetime comparison of DREEM with MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18] over 5 simulations (with 500 heterogenous nodes)

WSN#	MOCHA-15			HEEMP-18			ENEFC-17			DREEM		
	FND	HND	LND	FND	HND	LND	FND	HND	LND	FND	HND	LND
1	1149	1593	2182	2828	3344	3745	2271	2729	3842	3290	4195	4810
2	1092	1555	2227	2770	3525	3870	2196	2775	3737	3240	4221	4880
3	1091	1525	1989	2769	3410	3820	2256	2793	3671	3299	4207	4840
4	1185	1539	2174	2757	3559	3803	2191	2860	3926	3132	4199	4747
5	1144	1522	2142	2868	3577	3729	2165	2724	3631	3221	4190	4908
Average	1132.2	1546.8	2142.8	2798.4	3483	3793.4	2215.8	2776.2	3761.4	3236.4	4202.4	4837

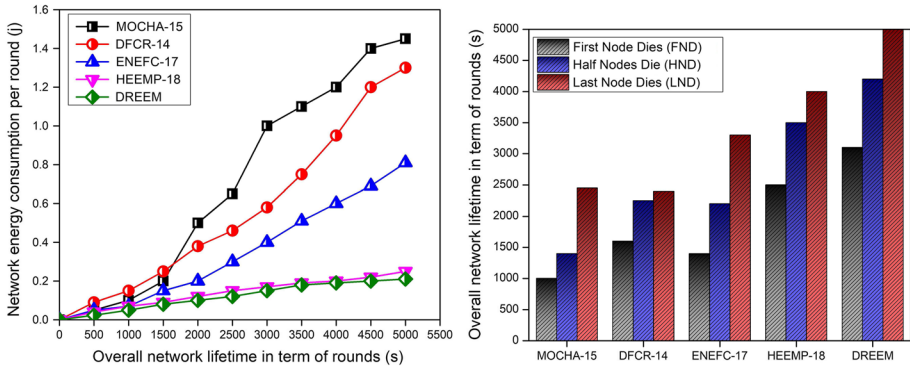**Fig. 4** The lifetime comparison of DREEM with recent schemes in the form stability period

Table 6 Dead node percentage comparison of DREEM with MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18] over 5 simulations (with 300 nodes)

% dead nodes	MOCHA-15	ENEFC-17	HEEMP-18	DREEM
10	1290	1330	2601	2806
20	1372	1475	2784	2951
30	1431	1610	2997	3331
40	1467	1727	3226	3597
50	1516	1898	3485	3959
60	1568	2034	3677	4311
70	1656	2176	3692	4588
80	1750	2296	3693	4733
90	1827	2437	3694	4891
100	2029	2960	3694	4809

Table 7 Dead node percentage comparison of DREEM with MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18] over 5 simulations (with 500 heterogenous nodes)

% dead nodes	MOCHA-15	ENEFC-17	HEEMP-18	DREEM
10	1295	1347	2565	2843
20	1325	1531	2652	2902
30	1366	1750	3208	3156
40	1452	1835	3299	3304
50	1520	1944	3591	3661
60	1556	2011	3680	3942
70	1672	2135	3672	4396
80	1681	2225	3792	4591
90	1796	2323	3753	4698
100	2126	3049	3704	4972

variance between their measures. All these tests are performed at the cluster head level and if the tests results are positive, then the cluster-head separates and deletes the redundant data from the available data sets. So, only limited data packets travel toward the BS those pass through the statistical tests, which decrease the overall network throughput as
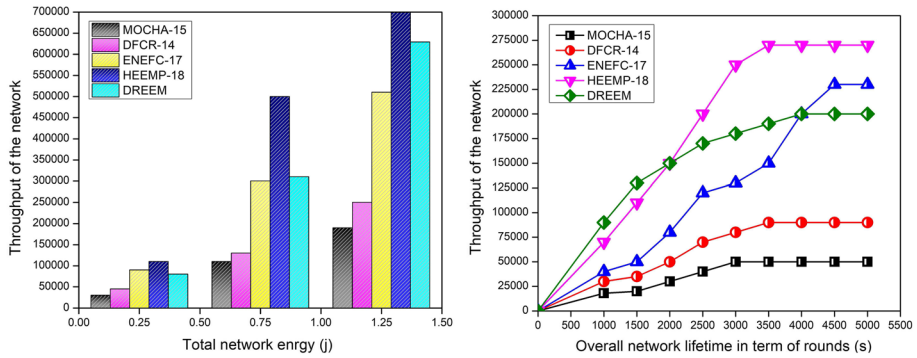
Fig. 5 The network throughput comparison of DREEM, MOCHA [17], DFCR [19], ENEFC [16], and HEEMP

compared ENEFC and HEEMP. As these models, MOCHA, DFCR, ENEFC, and HEEMP does not perform any similarity check for data-redundancy. Figure 5b portrays the performance of MOCHA, DFCR, ENEFC, HEEMP, and DREEM with different initial energies of field nodes. Tables 8 and 9 show the throughput comparison of MOCHA, DFCR, ENEFC, and HEEMP with node density $N = 300$ and $N = 500$. The DREEM remains consistent and performed favorably as compared to the MOCHA, DFCR, ENEFC, and HEEMP in all scenarios.

4.5 Average Number of CHs per Round

The field nodes are installed in the network sensing field through a distributed algorithm. As a result, the node density per unit area is not constant and in this way, the cluster head selected in the densely populated area has more member nodes. The poor CH selection criteria and CH mismanagement lead to more energy consumption. Figure 6a depicts the comparison of the number of CHs selected per round in DREEM, MOCHA, DFCR, ENEFC, and HEEMP. DREEM remains very consistent in selecting the CH with different node densities like $N = 100$ to $N = 500$ as depicted in Fig. 6b. The MOCHA CH selection criterion is poor, so, the number of CHs in MOCHA varies from $10 - 50\%$ CHs per round in comparison with others.

5 Conclusion

In this paper, we introduced a novel framework for data-aggregation in which the data-forwarding hierarchy is simple and selected routes are optimal with less network energy consumption. The proposed framework easily handles the limited energy challenge of WSNs by reducing the redundant and repeated transmissions over the link. The CH selection of our proposed framework dynamically rotates among the cluster member nodes, as a result, the nodes drifted anytime can be adjusted in a cluster as a member. The designed framework smartly decides the forwarders from the available forwarder list, and successfully eliminates the back-transmission problem. From simulation results, we note that the WSNs lifetime is totally reliant on the number of transmissions and redundancy can increase the

Table 8 Data packets comparison of DREEM with MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18] over 5 simulations (with 300 heterogenous nodes)

WSN#	MOCHA-15			HEEMP-18			ENEFC-17			DREEM		
	FND	HND	LND	FND	HND	LND	FND	HND	LND	FND	HND	LND
1	23224	32458	39856	42569	73288	88201	421940	689003	713256	295320	455255	623366
2	23560	34962	37526	45869	75821	89250	385623	585221	787633	302115	495900	618015
3	22861	29850	36325	44968	72587	87639	432567	652300	706651	299533	450035	659000
4	25293	33569	37814	43885	76356	86938	458965	665224	786350	315899	452224	631122
5	23510	28612	33215	47259	78233	90012	465663	652300	793600	325662	423698	644900
Average	23689.6	33890.2	36947.2	44910	75149	88408	432951.6	648809.6	757498	307705.8	455422.4	635280.6

Table 9 Data packets comparison of DREEM with MOCHA [17], DFCR [19], ENEFC [16], and HEEMP [18] over 5 simulations (with 500 heterogenous nodes)

| WSN#	MOCHA-15			HEEMP-18			ENEFC-17			DREEM		
	FND	HND	LND	FND	HND	LND	FND	HND	LND	FND	HND	LND
1	48522	66325	82569	46003	69233	96203	702361	818763	966996	369391	589214	912688
2	43652	63259	85960	49263	72623	98220	692000	802003	949863	311820	581477	946081
3	44255	65800	82666	47939	78520	97337	685666	821552	900201	378194	562149	921508
4	50026	63002	84002	40269	79623	96349	712119	848527	993452	358900	541899	991637
5	49028	69820	81032	45205	78590	98229	630022	858211	937786	332140	541633	970521
Average	47096.6	65641.2	83245.8	45735.8	75717.8	97267.6	684433.6	829811.2	949659.6	350089	567274.4	948487
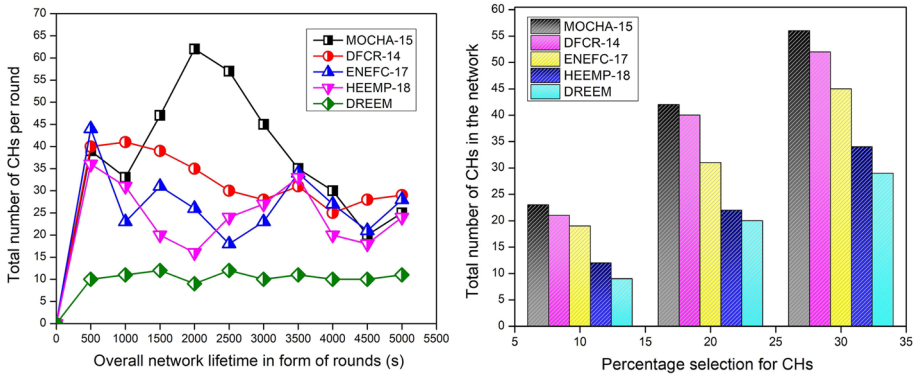
Fig. 6 The average number of CHs per round comparison of DREEM, MOCHA [17], DFCR [19], ENEFC [16], and HEEMP

energy consumption of the network. If we avoid the redundant transmission and control the auxiliary information not to circulate inside the network, we can save the $20 - 30\%$ energy consumption of the network. We perform simulation using five different metrics like the network lifetime, the energy consumption in stable and unstable network state, the network throughput and number of CHs in the network to validate our model. From the simulation results, we found that the DREEM has less energy consumption in comparison with current approaches. In future work, the lifetime of the DREEM can be further enhanced by taking into consideration the energy harvesting. We are also planning to study clustering with different renowned optimization algorithms and preparing to include the mobile CHs for data collection.

References

1. Karimi, H., Medhati, O., Zabolzadeh, H., Eftekhari, A., Rezaei, F., Dehno, S. B., et al. (2015). Implementing a reliable, fault tolerance and secure framework in the wireless sensor-actuator networks for events reporting. *Procedia Computer Science*, *73*, 384–394.
2. Ahmed, G., Zou, J. H., Fareed, M. M. S., & Zeeshan, M. (2016). Sleep-awake energy efficient distributed clustering algorithm for wireless sensor networks. *Computers and Electrical Engineering*, *56*, 385–398.
3. Chirihane, G., Zibouda, A., & Benmohammed, M. (2016). An adaptive clustering approach to dynamic load balancing and energy efficiency in wireless sensor networks. *Energy*, *114*, 647–662.
4. Ahmed, M., Salleh, M., & Ibrahim, M. (2017). Routing protocols based on node mobility for underwater wireless sensor network (UWSN): A survey. *Journal of Network and Computer Applications*, *78*, 242–252.
5. Khan, J. U., & Cho, H. S. (2015). A distributed data-gathering protocol using AUV in underwater sensor networks. *Sensors*, *15*(8), 19331–19350.
6. Javaid, N., Hafeez, T., Wadud, Z., Alrajeh, N., Alabed, M. S., & Guizani, N. (2017). Establishing a cooperation-based and void node avoiding energy-efficient underwater WSN for a cloud. *IEEE Access*, *5*, 11582–11593.

7. Bahi, J., Makhoul, A., & Medlej, M. (2014). A two tiers data aggregation scheme for periodic sensor networks. *Ad-Hoc & Sensor Wireless Networks*, *21*(1–2), 77–100.
8. Guangjie, H., Jiang, J., Bao, N., Wan, L., & Guizani, M. (2015). Routing protocols for underwater wireless sensor networks. *IEEE Communications Magazine*, *53*(11), 72–78.
9. Deqing, W., Ru, X., Xiaoyi, H., & Wei, S. (2016). Energy-efficient distributed compressed sensing data aggregation for cluster-based underwater acoustic sensor networks. *International Journal of Distributed Sensor Networks*, *2016*(19), 1–14.
10. Liao, Y., Qi, H., & Li, W. (2013). Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks. *IEEE Sensing Journal*, *13*, 1498–1506.
11. Dervis, K., Okdem, S., & Ozturk, C. (2012). Cluster-based wireless sensor network routing using artificial bee colony algorithm. *Wireless Network*, *18*, 847–860.
12. Orojloo, H., & Haghighat, A. T. (2015). A Tabu search based routing algorithm for wireless sensor networks. *Wireless Networks*, *22*(5), 1711–1724.
13. Tanwar, S., Tyagi, S., Kumar, N., & Obaidat, M. S. (2018). LA-MHR: Learning automata based multilevel heterogeneous routing for opportunistic shared spectrum access to enhance lifetime of WSN. *IEEE Systems Journal*, *13*(1), 313–323.
14. Lin, H., Chen, P., & Wang, L. (2015). Energy efficient clustering protocol for large-scale sensor networks. *IEEE Sensor Journal*, *15*(12), 7150–7160.
15. Fersi, G., Louati, W., & Jemaa, M. B. (2016). CLEVER: Cluster-based energy-aware virtual ring routing in randomly deployed wireless sensor networks. *Peer-to-Peer Networking and Applications*, *9*(4), 640–655.
16. Muthukumaran, K., Chitra, K., & Selvakumar, C. (2018). An energy efficient clustering scheme using multilevel routing for wireless sensor network. *Computers and Electrical Engineering*, *69*, 642–652.
17. Songhua, H., Jianghon, H., Wei, X., & Chen, Z. (2015). A multi-hop heterogeneous cluster-based optimization algorithm for wireless sensor networks. *Wireless Networks*, *21*(1), 57–65.
18. Sajwan, M., Devashish, G., & Sharma, A. K. (2018). Hybrid energy-efficient multi-path routing for wireless sensor networks. *Computers and Electrical Engineering*, *67*, 96–113.
19. Azharuddin, M., Pratyay, K., & Prasanta, K. P. (2015). Energy efficient fault tolerant clustering and routing algorithms for wireless sensor networks. *Computers and Electrical Engineering*, *41*, 177–190.
20. Vakily, T. V., & Jannati, M. J. (2010). A new method to improve performance of cooperative underwater acoustic wireless sensor networks via frequency controlled transmission based on length of data links. *Wireless Sensor Network*, *2*, 381–389.
21. Harb, H., Makhoul, A., Tawil, R., & Jaber, A. (2014). A suffix-based enhanced technique for data aggregation in periodic sensor networks. In *International wireless communications and mobile computing conference* (IWCMC), Nicosia, 494–499.
22. Tran, K. T. M., Oh, S. H., & Byun, J. Y. (2013). Well-suited similarity functions for data aggregation in cluster-based underwater wireless sensor networks. International Journal of Distributed Sensor Networks, *2013*, Article ID 645243,7.

Dr. Gulnaz Ahmed is currently pursuing post-doctorate from School of Management, Xi'an Jiaotong University, Xi'an, China. Previously, she completed Ph.D. from School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China. Her research interests include: image processing, big data analysis, wireless body area networks, and wireless sensor networks.
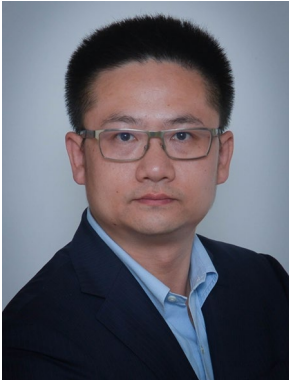
Xi Zhao (S'08–M'10) received the Ph.D. (Hons.) degree in computer science from Ecole Centrale de Lyon, Lyon, France, in 2010. He conducted research in the fields of biometrics, face analysis, and pattern recognition, as a Research Assistant Professor with the Department of Computer Science, University of Houston, USA. He is currently a Professor with Xi'an Jiaotong University, Xi'an, China. His current research interests include biometrics, affective computing, data analysis, mobile computing, computer vision, healthcare computing, and assistive technology. Dr. Zhao currently serves as a Reviewer of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON INFORMATION, FORENSICS AND SECURITY, Image and Vision Computing, the IEEE International Conference on Automatic Face and Gesture Recognition, the IEEE International Conference on Advanced Video and Signal-Based Surveillance, and the International Conference on Pattern Recognition. He has served as the Co-Chair of the International Conference on Biometrics: Theory, Applications and Systems in 2013, the Program Committee of Workshop on 3-D Face Biometrics in 2013, and the International Conference on Affective Computing and Intelligent Interaction in 2015.

Dr. Mian Muhammad Sadiq Fareed is currently pursuing post-doctorate at School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China. Previously, he completed Ph.D. in Information and Communication Engineering from Xi'an Jiaotong University, China. He is interested in computer vision, object detection/ segmentation, wireless body area networks, and wireless sensor networks.

Muhammad Rizwan Asif is currently pursuing Ph.D. in Information and Communication Engineering from Xi'an Jiaotong University, China. He is currently on leave from his position of Lecturer at COMSATS Institute of Information Technology, Lahore, Pakistan. His research interests include image processing, computer vision and intelligent systems.

Syed Ali Raza is enrolled in the Ph.D. in School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China. Previously, he completed MS Electrical Engineering School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China. His main research interests include wireless body area sensor networks, and wireless sensor networks.