Eng. Sci., 2022, **19**, 144–153

Engineered Science

DOI: https://dx.doi.org/10.30919/es8d694

Energy Efficient Data Aggregation and Density-Based Spatial Clustering of Applications with Noise for Activity Monitoring in Wireless Sensor Networks

Jainendra Singh[1,*] and Zaheeruddin[2]

Abstract

Clustering saves energy and extends network longevity in applications such as activity monitoring in wireless sensor networks. Clustering separate tes nodes into clusters from which the cluster head can collect and send data to the base station. The current investigation also demonstrates data aggregation at the cluster head before sending it to the base station. The cluster head is chosen based on leftover energy and distance from the sink and other network nodes. The current study found that the existing clustering protocol decays its energy quickly, but our proposal can overcome these challenges and sensors can survive longer. This paper proposes Density-based Spatial Clustering of Applications with Noise (DB-SCAN) clustering algorithms to make data gathering and transmitting of data energy efficient. The data aggregation protocol will collect the packets and combine them. Further, it will be forwarded using the clustering approach. It is used to improve transmission efficiency by reducing the transmission of redundant data to the sink. This proposed protocol, Energy Efficient Cluster based Data Aggregation (EECDA), will help increase network lifetime by ensuring uniform distribution of energy amongst the nodes.

Keywords: Wireless sensor network; Clustering; Spatial clustering; Data aggregation; Activity monitoring; Energy.
Received: 25 February 2022; Revised:12 April 2022; Accepted: 13 April 2022.
Article type: Research article.

1. Introduction

Wireless sensor networks (WSNs) are a group of sensors that are dispersed around the region of interest (ROI). Each sensor node can provide data to the transceiver to detect certain phenomena through sensing, such as automobile tracking, wildlife tracking, environmental monitoring, military, and so on.

The job of the transceiver is to collect data from the sensor node using the internet as shown in Fig. 1. The data transmission, the transceiver connects to the network. The data from the surroundings is stored in the sensor node and forwarded to the base station via multi-hop communication. WSNs are data-gathering networks made up of small sensors dispersed around an area of interest. The application scenario in WSNs is often intervention-free and difficult due to geographical limits. WSNs use small sensors to collect data from their surroundings and transfer it to a sink. To forward data, single-hop or multi-hop routing can be utilized. The network topology is set up such that neighboring nodes can choose a cluster head (static or dynamic) to assist with data forwarding based on the criteria. There are numerous clustering and cluster head selection options for the operation of WSNs.[1]

In light of the popularity of the internet of things in a variety of applications such as healthcare, forest fire detection, defense, and home automation, WSNs are among the most promising technologies to emerge in recent years. The fundamental reason for this increase is the many benefits of adopting WSNs, including their cost-effectiveness, small size, scalability, and other characteristics.[2] The majority of WSNs are built of sensor nodes (SNs), which are randomly dispersed around the region of interest. There are sensors in each node, and each sensor node is equipped with a sensor module, a transmitter, a receiver, a processing unit, and a buffer, and all of these are powered by battery power. The cells used in the SNs have a limited capacity, and replacing and recharging them is a time-consuming job. As a result, it uses more energy as a result of the massive data transfers than doing other SN tasks. However, data transmission is required for WSNs

---

[1] *Department of Electrical Engineering, Jamia Millia Islamia, New Delhi, India, 110001.*

[2] *Department of Electrical Engineering, Jamia Millia Islamia, New Delhi, India, 110001, Centre.*

[#] *These authors contributed to this work equally.*

**E-mail:* er.jainendra@gmail.com (J. Singh)

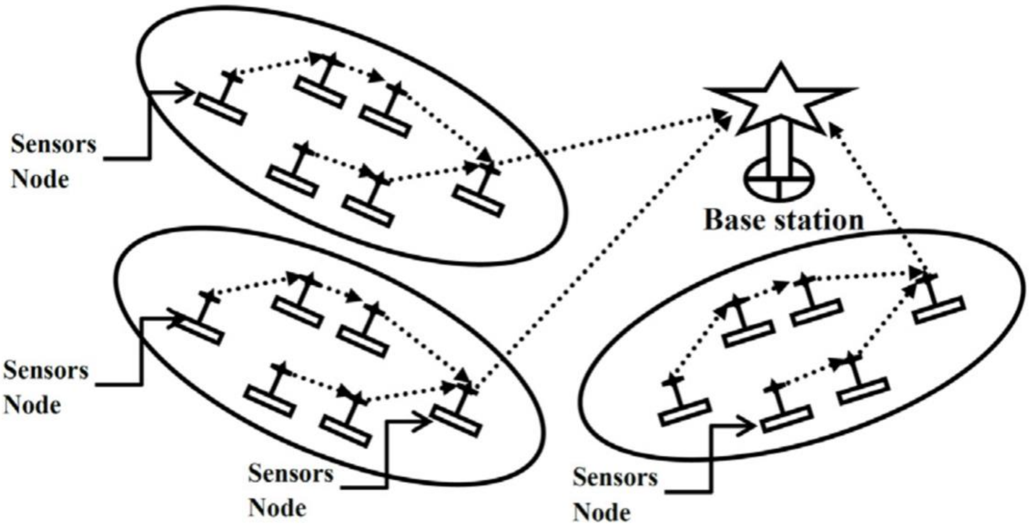

Fig. 1 Clustering and Data Aggregation operation overview in wireless sensor networks.

because the central station, referred to as the sink or base station (BS), uses the data to perform additional operations on the network. These data transmissions are primarily accomplished through the use of a multi-hop transmission medium, in which some of the SNs are responsible for transferring the data to the sink. Because these relay nodes are involved in more data exchanges and experience an earlier battery drain, they are more prone to problems.[3] It is difficult to solve because it isolates a portion of the network and causes the data collection process to be interrupted. By visiting a sink node (SN) in the WSN, a mobile sink (MS) may help to ease the hotspot issue. However, it is not feasible and results in a number of difficulties such as delay, significant packet loss, congestion, and so on. A set of visiting locations known as rendezvous points (RPs) is identified in WSN, rather than visiting each node. The MS only visits the RPs, and other devices send their packet data to the nearest accessible RP instead of visiting each node. On the other hand, all of this research is concentrated on a network that generates data in a uniform manner. The WSN applications, including forest fire detection, healthcare delivery, homeland security, and home automation, are event-driven in their design and functionality. Because the data is generated only when an event occurs in the network, event-driven applications are more efficient.[4-6] But they keep a constant eye on the surrounding area. Data is transferred to the sink node once an event occurs in the point of interest and only when this event occurs. This results in certain SNs producing large amounts of data on a regular basis, while others produce just a small quantity of data on a periodic basis.

The classification suggested in recent studies[7] states that WSNs have mainly three critical parameters, energy efficiency, delay, and QoS. The medium access control (MAC) protocol plays a vital role in energy consumption because it controls the duty cycle of the radio.[8-10] In communication, the several challenges[7] need to be addressed before protocol design. These protocols have been extensively studied in Ref.[8,11-13] such as dataset has been suggested which can be very useful for simulation validation. The internet of things (IoT) represents the advancement of digitization in our lives, society, and economy. Through the use of an amazing range of applications, large-scale intercommunication is possible between objects and people.

In our study, we make the assumption that in WSNs the cluster head collects data and forward using multi-hop communication. We propose Density-based Spatial Clustering Applications with Noise (DB-SCAN) for WSNs; it provides path planning using a Steiner point around those cluster heads' communication range. The mobile sink follows the data collection path, which seems more efficient than the existing proposals on many parameters.[14-16] This section discusses the various factors that motivate the selection of this project for our dissertation. Firstly, IoT is swiftly and steadily becoming a major part of our lives, be it in the development of smart cities, smart home appliances, vehicles, environmental monitoring to control issues of global warming, and pollution, and in disaster management. Hence, efficient data aggregation becomes an important criterion to be considered for effective communication and fast delivery of services. Secondly, the massive amount of data that is generated and transferred within networks contains a prominent portion of data redundancy which leads to decreased network lifetime. Therefore, we are motivated to devise one such technique such that only the required data is sent to the sink node. The rest of the paper has been organized in the following way as in Section 2 the methodology has explained associated with the method of data collection using the mobile sink. In Section 3, our proposal's results and evaluation are given, followed by the conclusion in Section 4. The background provides an insight into how WSNs and IoT systems have evolved through time and they continue to do so. The present study presents the objectives for the work and how it aims to make significant contributions to this novel model.

A survey has been provided by Ref.[12] which explains the limitations of traditional networking in handling IoT requirements. In addition, they provide the guideline to achieve adequate control and access to connected devices in a flexible, scalable, efficient, seamless, and cost-effective manner. The software-defined networking (SDN) based technologies come to achieve the above-mentioned objective. The software-defined internet of things (SDIoT) is the new networking paradigm that has four major aspects: edge networking, access networking, core networking, and data networking. Data aggregation is an important aspect of SDN-based edge networking. Energy conservation is the key issue in WSNs since a large number of deployed sensor nodes are likely to transfer redundant data which increases communication cost and in turn decreases network lifetime. Energy conservation, query processing, and communication are major topics to be focused on. The big data paradigm and the integration of its major principles in the WSN technology are introduced by Ref.[5] The four-layered WSN framework for the IoTs ecosystem developed in Ref.[6] has provided the specifications. The different layers of communication are sensors, nodes, cluster hubs, and cloud servers. This framework has been tested for healthcare applications and can be further implemented in diverse IoT applications. Castro demonstrated an IoT-based WSN for monitoring environmental temperature and humidity, which was then used for product maintenance and pharmaceutical organizations. Following good feedback, expanding this framework across the whole hospital has become a viable option.

The data collection approach using mobile sink can be classified into two categories:

(1) Direct and

(2) Rendezvous point (RP).[5]

Simple medium access control (SMAC)[15] is an early-developed contention-based MAC protocol. One of the fundamental schedule-based MAC protocols is low energy adaptive clustering hierarchy with deterministic cluster-head selection (LEACH),[16] a cluster-based scheduling protocol. The detailed description of scheduling in WSNs is documented in Ref.[17] and it will focus our related work on the mobile sink in WSNs only. "To sleep or not to sleep" was the question driving the dynamic sensor MAC protocol (DSMAC). While the static low-duty cycle assignment in SMAC has saved energy, the latency issue has been disregarded. Latency is a critical factor in the usability of sensor networks in time-sensitive applications like healthcare and the military. Depending on the application, the duty cycle may be manually changed. This dynamic SMAC protocol may automatically alter its duty cycle without prior application knowledge. Based on these observations, they developed the dynamic sensor MAC protocol. This allows DSMAC to change the sleeping interval with a fixed listen-to interval length. As a consequence, DSMAC eliminates SMAC's significant delay problem while preserving energy efficiency. DSMAC also has less overhead than SMAC.[15-18] Tree-based LEACH protocol named T-LEACH[19] requires cluster leaders to rotate every batch of rounds rather than every round. Nodes continue to lead clusters if their energy surpasses a particular level. The modified tree cluster head rotation (MT-CHR) method overcomes the T-LEACH protocol's major weaknesses. A novel MT-CHR[20] protocol threshold energy expression is also introduced to delay initial node death and avoid data loss. MT-CHR performance is measured by alive nodes, network longevity, and network use. The results are compared to LEACH and T-LEACH results, and the MT-CHR technique contributes significantly. It is incredibly versatile and effective in assuring long-term sensor networks. An improved approach for determining the distribution function of multi-hop latency measurements is presented in this paper. This innovative method was motivated by the distribution of broadcast delay as a function of node geometry (grid topologies). The proposed multi-hop computation may utilize the provided one-hop delay distribution calculation. This study's notion allows the estimation of probability distribution functions without simulation[21] or iterative numerical techniques.[22] As the distance from the transmitting source increases, so do the variations in dispersion. As a consequence, the underestimation of latency in broadcast transmissions may be prevented, which is unachievable with earlier approaches that simply use the mean latency computation. Protocols like route discovery and route maintenance have precise timing values; approximated values are unacceptable for delay-constrained applications. A novel routing strategy[23] for energy optimization based on LEACH for WSN is provided as the roadmap. The fundamental difficulty with WSN routing is network longevity. They designed a recommended routing protocol that blends micro genetics and LEACH to solve this problem. Compared to existing protocols, the existing GA-LEACH protocol enhances CH selection while reducing network energy utilization. This study compares the findings to known hierarchical routing protocols such as LEACH, LEACH-C, LEACH GA, and GADA LEACH routing protocols with varying packet sizes and beginning energy.

Compressive sensing (CS) is used to obtain data from WSNs.[24] CS retrieves all WSN data at the sink node. With cluster-based techniques, the CS algorithm saves energy. This is based on M measurements and N nodes. The Fourier transform guarantees excellent stability with little processing. This promotes orthogonal matching pursuit. We compare ACS-LEACH to contemporary NS2-based WSN systems. Achieved in 90-node topologies are 0.018, 0.012, and 0.0011 bits/s.

In Liang 2019,[25] several sensor devices are placed throughout an area in a wireless sensor network to gather for analysis. They are scattered throughout and have a battery. The battery is also small owing to its size. Its main purpose is to modify the original LEACH technology to save energy named as modified medium head in leach (MMH LEACH). The approaches outperform the original LEACH algorithm in residual energy. When used in a bigger setting, network life increases. The LEACH technique saves energy. The network's longevity is also important. WSN protocols.[26] REACT-LEACH is a RE-aware clustering transformation approach for LEACH that incorporates a clustering mechanism. Because REACT-LEACH measures residual energy, the recommended cluster head rotation, and cluster reformation approaches work better. In addition, to maintain the readability of the presented text the abbreviations has given in Table1.

Table 1. Abbreviations used in presented work.

CH	Cluster head	LEACH	Low energy adaptive clustering hierarchy
RE	Residual energy	TDMA	Time division multiple access
DA	Data aggregation	MAC	Medium access control
CF	Cluster formation	DSMAC	Dynamic Sensor MAC protocol
TE	Total energy	*ECU-dist*	Euclidean distance
PF	Probability function	SDIoT	Software-defined internet of things
PFO	Path formation	ACS	Adaptive Compressed Sensing.

2. Methodology

The sensors and network model are based on the following assumptions the sink node is situated in the center of the sensing area. The location-aware item sensors are energy-constrained therefore energy-aware and resource-adaptive item sensors are available. In addition, a symmetric propagation channel with stationary nodes and a sink is also available. On the transmitter and receiver sides, energy usage can be measured. Assume that $E_{tx}(left)(l,d,right)$ and $E_{rx}(left)(l,right)$ is the energy consumed by the transmitter and receiver units, respectively, while broadcasting a 1-bit message, as calculated by Eqs. (1-2). A node's transceiver circuitry is given as:

$$E_{tx}(l,d) = E_{Cct}(l) + E_{amp}(l,d) \tag{1}$$

$$E_{rx}(l) = E_{Cct}(l) = l \times E_{Cct} \tag{2}$$

The selection of $E_{tx}(l,d)$ model at instance t depends on the distance d between the transmitter node p_i and receiver node p_j and is given by Eq. (3).

$$\begin{aligned} E_{tx}(l,d) = \{l \times E_{Cct} + l \times \varepsilon_{fs} \times d^2\ d \\ < d_0\ l \times E_{Cct} + l \times \varepsilon_{mp} \times d^4, \\ d \geq d_0 \end{aligned} \tag{3}$$

where, ε_{fs} and ε_{mp} represent free space and multi-path fading channel models, respectively. If d is greater than a given threshold $d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$ then the multi-path fading channel model is utilized. Otherwise, the free-space model is used. As a starting point, the following research makes use of the energy consumption and radio model that was previously discussed. When transmitting a k-bit message over a long distance, it is critical to maintaining an optimum signal-to-noise ratio (SNR) and distance d the energy cost of transmission (ET_X) is given in the later part of the text.

2.1. Proposed work

In this section, the prior requisite, which plays an essential role in our proposal is divided into two separate phases setup and steady. In the first phase, the nodes which are deployed in the region initialize and start the neighbor discovery.

It has set parameters; nodes select their leader node, which will be responsible for forwarding the data using multi-hop communication. This phase is called the network setup phase. There are plenty of centralizing, and distributed methods have been proposed for the selection of leader nodes. LEACH works by randomly selecting cluster heads, which collect data from sensors in their vicinity and transfer it to the base station. Since data transmission dissipates energy, cluster heads are re-selected after each round. The operation of the LEACH protocol consists of several rounds, each of which is treated as a unit, and every round consists of two phases named as setup phase and the steady phase. The following are the stages of the LEACH protocol: initial setup during the setup phase, a pre-determined group of nodes, p, is chosen as cluster-heads using the following Eq. (4). Every sensor node n selects a number r at random from 0 to 1.

$$p(n) = \frac{p}{1-p*(r*modp^{-})} \quad \forall n \in G \tag{4}$$

The r is a round number in this case. The chance of being chosen as CH is $frac1p$, where G is the set of nodes that is not selected as cluster head in earlier rounds. The threshold is T(n). Non-CH nodes get the cluster head advertising in the second phase and then submit a join request to the cluster head, indicating that they have joined the group underneath that base station. In the third phase, each of the selected CHs generates a TDMA dynamically updated for their cluster's member nodes. The phase of constancy cluster nodes transmits data to the central in a constant phase.

A one-hop transmission is a sole way for each cluster's member sensors to communicate with the cluster's main sensor. Messages are transmitted between the cluster head and the central node. After a given amount of time has passed, the network returns to the initialization and begins a new round of CH selection. The stochastic cluster-head selection in LEACH has the potential to create an energy imbalance in nodes, increasing the system's total energy waste. To ensure an equitable energy load distribution across the network, we add new variables depending on the RE. Our method's major aim is to avoid choosing CHs for nodes that have less residual energy and are farther away from the sink.

$$D_{net}(i) = \sum_0^i n\, D_{ij} + D_{iSink} \tag{5}$$

It will sort the D_{net} in ascending order to find the nodes with the highest eligibility of being CH as Eq. (5). The nodes which have larger RE are in AE. Also, to ensure that the nodes with similar above parameters can be selected CHs fairly, a randomness factor is introduced. We use the neighborhood concept to form the cluster, according to which a given node becomes a cluster member under a given CH provided that the CH is the closest node from the set of CHs which lie in its radio transmission range.

2.2 Cluster formation and data aggregation using modified DB-SCAN

It has been observed that cluster formation and data aggregation using modified DB-SCAN in this section combined. The DB-SCAN algorithm has the key playing role of PF. In addition, the CH which has α in its leading role will be iterative. It has been shown that the construction of clusters and the aggregation of data are performed via the use of a modified DB-SCAN algorithm. Assuming the worst-case scenario, the DB-SCAN algorithm, which iterates over the data, is crucial to the performance of the PF since it is responsible for the PF's overall performance. Aside from that, the CH who occupies a critical position in the iterative process will also be iterative. The initial detail is shown in Eq. (6).

$$\Gamma_{ij}^{a} = (1-\alpha)\Gamma_{ij} + \sum_{a=1}^{m} n\, \Delta\Gamma_{ij}^{a} \tag{6}$$

where α is the evaporation rate, S_i to S_j is the distance between the CH. The $\Delta\Gamma_{ij}^{a}$ will be calculated using Eq. (7), where the m will be the cluster head.

$$\Delta\Gamma_{ij}^{a} = \{Q \times L_a^{-1}\ if\ a\ travel\ between\ S_i\ and\ S_j, \quad Otherwise \quad 0 \tag{7}$$

The trajectory length will be represented by L_a. The Q is a constant. First, the probability function Eq. (8) will be estimated for each cluster head. The reasoning behind the probability function is that it can later help to increase the network lifetime.

$$p_i^{a} = \{\frac{\Gamma_{ij}^{\alpha} W_i^{\beta} \eta_{ij}^{\gamma} \xi_i^{\delta}}{\sum_{j\in N(i)}^{i} n\, \tau_{ij}^{\alpha} W_i^{\beta} \eta_{ij}^{\gamma} \xi_i^{\delta}} \quad iff\ N(i) \neq \phi, \quad Otherwise\ 0 \tag{8}$$

where α, γ, τ, β and η are absolute importance. Whereas, these values considering the clustering will be greater than 0. Eq. (8) and Eq. (10) ensure that it's an iterative process. Once the expected (L_m) achieved that the iteration will stop. In the meantime, all the selected CHs will be validated to check whether the parameters are being optimized. In addition, all the constraints such as W and γ also being fulfilled need to be validated. The weight variable is dynamic and therefore can tune in order to achieve the optimization objective. The α, β, γ, and δ entities should lie before 1 and greater than 0. Eqs. (6)-(8) will be used to estimate τ, W, and ξ, and Eq. (9) is used to estimate η.

$$\eta_{ij} = d_{ij}^{-1} \tag{9}$$

In Eq. (9), the measurement of SNs S_i, and selected CH S_j is stored in d_{ij}. The PF is a function that allows us to choose the node as CH which are having maximum occurrence or a maximal neighborhood. In each iteration, the process continues, and in the end, when no nodes are left to participate in the process, the process stops. In this way, the CH will be selected and in between the path will also start forming. This is one of the optimized ways of CH selection and PFO. In addition, in this work different PF is being used as it is shown in Eq. (10).

$$\psi_{ik}^{a} = \{\frac{\tau_{ik}^{\epsilon} \eta_{ik}^{\varrho}}{\sum_{k\in M(i)}^{n} n\, \tau_{ij}^{\epsilon} \eta_{ij}^{\varrho}} \quad iff \quad M(i) \neq \phi, \quad Otherwise\ 0 \tag{10}$$

where ϵ, τ, ϱ and η are the relative importance of the objective function ψ_{ik}^{a}. These values considering the clustering will be greater than 0. Once the expected (ψ_{ik}^{a}) is achieved so that the iteration will be stopped. The weight variable is dynamic and therefore can tune in order to achieve the optimization objective.

3. Results

In this section, the present study will discuss the results of our simulation. The simulation is performed using the NetworkX package in the Python and Matlab program. The simulation parameters have been given in Table 2. The present study has also assumed that all nodes have direct communication with the sink node. Nodes and sink items are not allowed to move around. In the first round, all nodes start with the same amount of energy. The energy used to transmit a packet from node 1 to node 2 is the same as the energy used to transmit a packet from node 2 to node 1. We assume that the sink has a static place in ROI and theoretically infinite resources in terms of energy. In addition, the node dies when its energy value reaches zero.

Table 2. Simulation parameter NetworkX package based on Python simulator

Number of nodes (n)	100
Area	(0,0) to (30,30)
p	0.1
sink. x, sink. y	(15,15)
$E_0(J.node^-1)$	0.5
$E_{TX}(nJ.bit^{-1})$	50
$E_{RX}(nJ.bit^{-1})$	50
Gainsboro! 60 EDA (nJ)	5
$E_{fs}(nJ \cdot bit^{-1}.m^-2)$	10
$E_{mp}(pJ \cdot bit1^{-1}.m^{-4})$	0.0013
r_{max}	500
Node Distribution	Random
DpacketLen (bits)	4000
Gainsboro! 60 HpacketLen (bits)	100

In Fig. 2, the observation of about 100 nodes scattered in 30 m^2. The base station is situated in the heart of the city. The nodes are thought to be stationary. Fig. S1 depicts the network model at the simulation's first phase when there are no dead nodes. The image depicts the entire process of CH selection

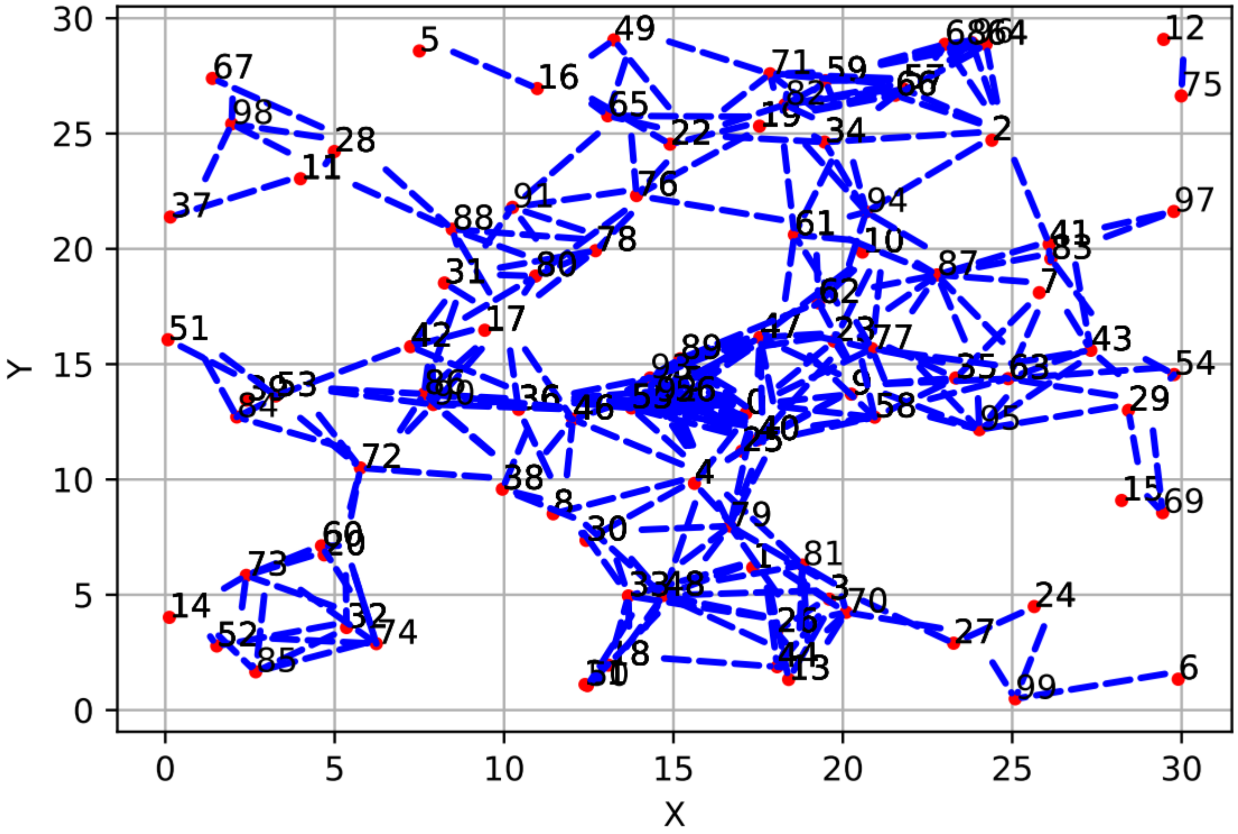

Fig. 2 Node Deployments in the area of interest for activity monitoring.

and CF in the area of interest of size (X*Y) m^2. Blue lines reflect data sent from cluster nodes toward CH. The CH performs DA before sending the aggregated data to the sink. Clustering is done in our proposed algorithm by considering the RE parameter, in addition to the distance between each node and all other nodes, as well as the distance between each node and the sink, resulting in a consistent energy distribution and efficient CH deciding factor and CF. In Fig. 2, nodes farther from the sink form clusters and sent their data in direction of its CH. But, nodes near the sink without any hurdle forward its data as less energy would be required to send data to the sink. The TE of the model and the energy of individual nodes significantly decrease in this phase and when the energy of the node becomes less than zero then the node dies.

Fig. 5d displays the EC of the nodes per round in the two protocols. In the graph, the total energy consumed in the case of LEACH increases at a faster rate as compared to EECDA. EECDA considers RE to calculate the next CH that improves the utilization of energy compared to the dead nodes denoted by red dots.

In the case of LEACH, the instability period is very high as the difference between the time when the first node dies and the entire network dies is very large. But in the case of our proposed algorithm, the instability period is low due to load balancing in the network. This is very important because in many applications when 50% of the nodes in a network die then the entire network becomes unserviceable.

In Fig. 5a, the number of alive nodes per round is compared between LEACH and EECDA. The number of live nodes directly affects the network's lifetime and stability of the network. In LEACH, nodes begin to die after the completion of 130 rounds whereas in the case of EECDA dead nodes start appearing after 220 rounds. All the nodes in both protocols die at around 250 rounds. The instability period which is the difference between the time when all nodes are dead and the initial node is inactive is higher in the case of LEACH than in the case of EECDA. This result shows that the proposed algorithm has an advantage when it comes to network lifetime and stability of the network compared to LEACH. The energy is also evenly scattered with the network as the RE of a node is considered during CH selection instead of randomly selecting cluster heads. The distance of a node from the sink is also taken into account during cluster formation. In Fig. 5a, the average energy of sensors per round is compared between LEACH and EECDA. The average energy of the sensor is the effective energy per sensor in a particular round. As can be observed in Fig. 5b, in the case of LEACH protocols, nodes start to die at around round 130 and at around 220 in the case of EECDA protocol. The total energy of nodes in the case of LEACH also decreases at a faster rate than in the case of EECDA. All nodes die at around 250 due to which average energy becomes 0. After round 200, there is a period when the average energy of sensors in the case of LEACH is more than in the case of EECDA.

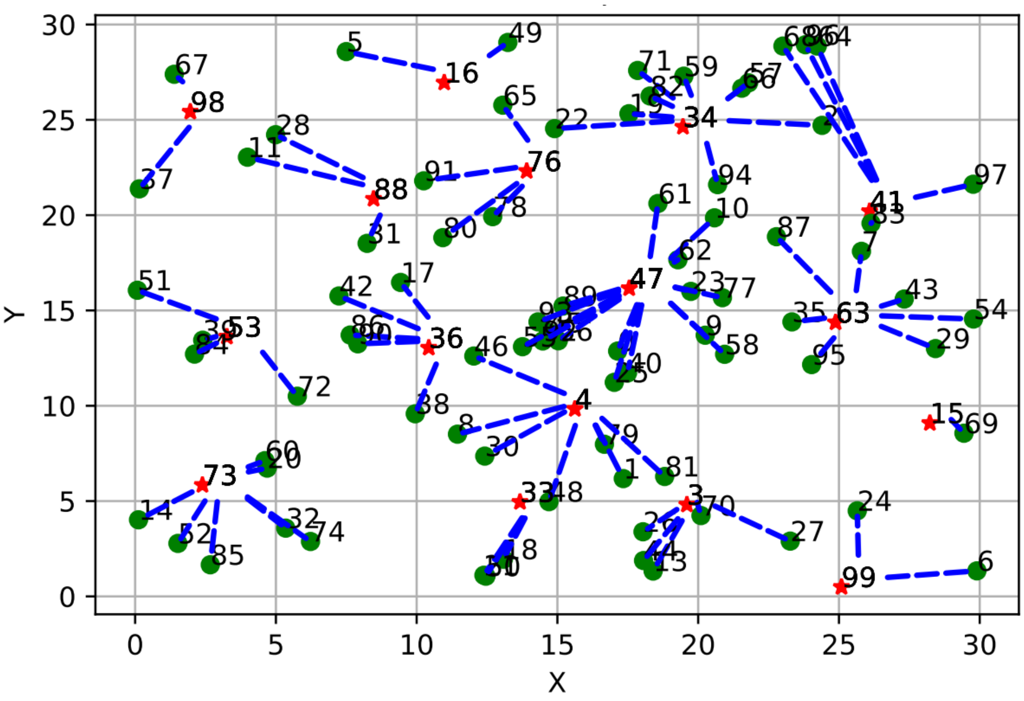

Fig. 3 Cluster formation and cluster head selection from DBSCAN clustering.

This occurs as a few nodes with high energy remain alive in LEACH till the end due to non-uniform energy distribution. But in the case of EECDA, there is uniform energy distribution due to which most of the nodes die in a very small time period. LEACH selects CH randomly. Figs. 2-4 shows the network status once the clustering protocol initializes formation and selects the cluster head. In our case, the present study is performing a DB-SCAN clustering approach, which can be replaced with any clustering method. The present study can observe the network status without the range disk f cluster head in Fig. 3. The most common path formation practice is to use the head node position to collect the data. The closed-loop uses a common approach to form the path considering the entire head node shown in Fig. 4. In the present study, the existing disk graph property of the head node to find the most suitable Steiner point as described in AlgorithmS2. Data aggregation at the cluster head reduces the transmission cost and reduces total EC. In LEACH, the CH has packets from nodes

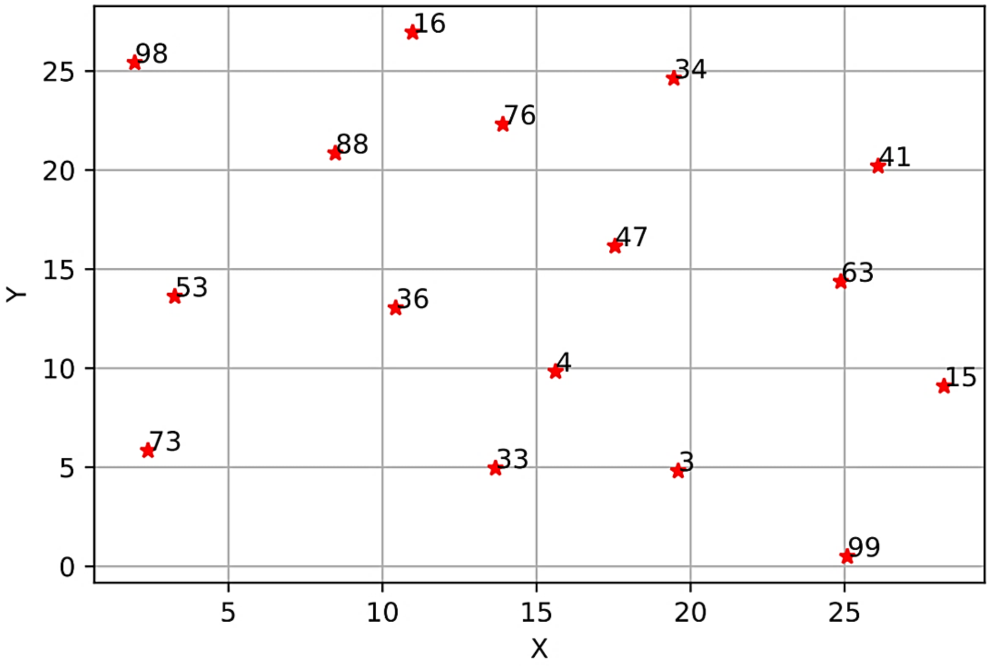

Fig. 4 Formation of dead nodes (the final phase).

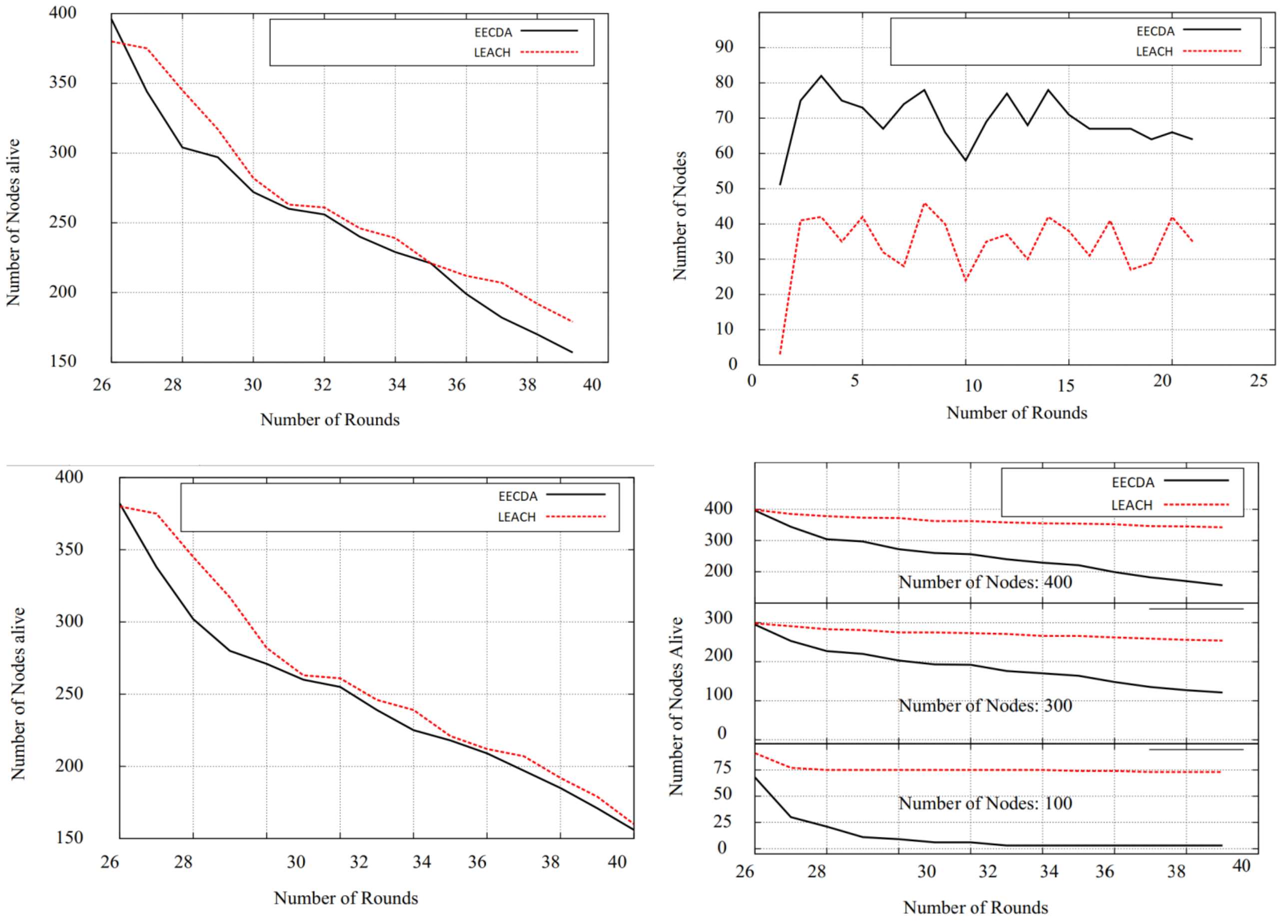

Fig. 5 Simulation results based on the proposal for parameters such as delay, energy, and throughput (a) Number of alive sensors after round, (b) Average energy of sensor after round, (c) Total energy consumed by the sensor after round, and (d) Number of packets sent to sink per round.

and sends the packets to sink in every round. But in the case of EECDA, redundant data is removed using data aggregation and only distinct data is sent to the sink in every round. In Fig. 5a, the AE of sensors in the context with |nodes| is plotted and the results are compared between LEACH and EECDA. In the proposed Algorithm S1, the CH is chosen considering the residual energy of sensor nodes due to which energy is distributed uniformly in the network. The graph of AE also decreases uniformly with an increase in the number of nodes. In the case of LEACH, due to random CH selection and cluster formation, energy is not distributed uniformly and varies largely with the difference in the |nodes|. Fig. 5a also displays that the AE of the algorithm is always more than that of LEACH due to efficient clustering and data aggregation used in Fig. 5d, the |data packets| sent to the main node per round are compared between LEACH and EECDA. As shown in Fig. 5d, the dynamic of data traffic is playing a vital role in performance. As the number of packets varies over the period, the LEACH is having a hard time stabilizing. On the other hand, our proposal is able to perform well even though the data packets vary frequently. Our approach's advantage is that it minimizes travel time by opting for the RP far from the head node but near enough to collect data. In this way, it reduces the mobile sink's travel distance to any almost multiple of a head node with its disk radius value. It is a significant gain over the existing approach. The virtual division of a single position into four different places gives the locality-based path direction profitable for our cost optimization.

4. Conclusion

The paper proposes a cluster-based data aggregation protocol that can increase the duration of the sensor network which optimizes cluster head selection with various parameter considering. Clustering is one of the energy-efficient techniques for extending the network lifetime. Clustering techniques organize nodes into a cluster where the CH can collect data from nodes in their respective cluster and transfer it to the base station. Further, the present study also introduces a data aggregation technique that aggregates the data at the CH and then transfers the aggregated data to the base station. The present study compares the proposed method with the LEACH protocol. Nodes that have previously been cluster heads in LEACH are unable to become cluster heads for P rounds, where P is the required proportion of cluster heads. The cluster head is chosen by EECDA based on the nodes' leftover energy and their distance from the sink and other parts of the network.

Sensors in the LEACH protocol begin to deteriorate in the early rounds, but sensors in the EECDA protocol stay alive for longer. LEACH's energy usage grows at a quicker pace, reducing the total network lifetime. Because it combines the information on the base station, the number of packets transmitted in leach is higher than in EECDA. As the number of nodes in a network grows, LEACH acts inefficiently, with huge fluctuations in average sensor energy, resulting in network instability, whereas EECDA maintains a consistent energy drop.

Acknowledgments

The work is independently done by authors and involved parties acknowledge in reference.

Conflict of interest

The authors declare no conflict of interest.

Supporting information

Not applicable.

References

[1] A. Chakrabarti, A. Sabharwal, B. Aazhang, *Information Processing in Sensor Networks*, 2003, **20**, 129–145, doi: 10.1007/3-540-36.

[2] H. Ahmad, C. Shivalingagowda, N. Kohli,. Kumar Sah, Coverage Optimization using Nature-Inspired Algorithm for Directional Sensor Networks, *Nature-Inspired Computing for Smart Application Design* , 2021, 169-191.

[3] D. K. Sah and A. Tarachand, *Computer Science Review*, 2018, **27**, 112–134, doi: 10.1016/j.cosrev.2017.12.002.

[4] A. Chakrabarti, A. Sabharwal, B. Aazhang, *ACM Transactions on Sensor Networks*, 2006, **2**, 297-324, doi: 10.1145/1167935.1167936.

[5] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand, A. H. Gandomi, *IEEE Internet of Things Journal*, 2019, **6**, 5132-5139, doi: 10.1109/jiot.2019.2897119.

[6] S. Chaya, P. V. Y. Jayasree, S. Kumar, D. K. Sah, Boolean directional sensor orientation solution for K-coverage in wireless sensor network, *4th IEEE International Conference on Recent Advances in Information Technology (RAIT)*, 2018, doi: 10.1109/RAIT.2018.8389090.

[7] C. Shivalingagowda, H. Ahmad, P. V. Y. Jayasree, D. K. Sah, Wireless Sensor Network Routing Protocols Using Machine Learning, *Architectural Wireless Networks Solutions and Security Issues*, 2021, doi: 10.1007/978-981-16-0386-0_7.

[8] C. Shivalingagowda, P. V. Y. Jayasree, D. K. Sah, Efficient energy and position aware routing protocol for wireless sensor networks, *KSII Transactions on Internet and Information Systems*, 2020, **14**, 1929–1950, doi: 10.3837/tiis.2020.05.004.

[10] D. K. Sah, K. Cengiz, P. K. Donta, V. N. Inukollu, T. Amgoth, *Computer Communications*, 2021, **180**, 48-56, doi: 10.1016/j.comcom.2021.08.017.

[10] D.K. Sah, K. Cengiz, P. K. Donta, V. N. Inukollu, T. Amgoth, *Computer Communications*, 2021, **12**, 180-192, doi: 10.1016/j.comcom.2021.08.017.

[11] D. K. Sah, T. N. Nguyen, K. Cengiz, B. Dumba, V. Kumar, *Cluster Computing*, 2022, **25**, 1715-1727, doi: 10.1007/s10586-021-03316-1.

[12] D. K. Sah, C. Shivalingagowda, D Praveen Kumar, Optimization problems in wireless sensors networks, *Soft computing in wireless sensor networks*, 2018, 29-50, doi: 10.1109/CISIS.2011.50.

[13] G. Xing, T. Wang, W. Jia, M. Li, Rendezvous design algorithms for wireless sensor networks with a mobile base station, *9th ACM international symposium on Mobile ad hoc networking and computing, ACM*, 2008, **13**, 231–240, doi: 10.1145/1374618.1374650.

[14] S. Karimullah, D. Vishnuvardhan, K. Riyazuddin, K. Prathyusha, K. Sonia, Low power enhanced leach protocol to extend WSN lifespan. *Lecture Notes in Electrical Engineering*. 2020, 527-535, doi: 10.1007/978-981-15-7961-5_51.

[15] K. Amandeep, R. Kumar, A Comparative Analysis of Improvements in Leach Protocol: A Survey, *Mobile Radio Communications and 5G Networks*, 2021, **5**, 129-143, doi: 10.1007/978-981-15-7130-5_9.

[16] H. Liang, S. Yang, L. Li, J. Gao, Research on routing optimization of WSNs based on improved LEACH protocol, *EURASIP Journal on Wireless Communications and Networking*, 2019, **11**, 1-12, doi: 10.1186/s13638-019-1509-y.

[17] M. J. Handy, D. Timmermann, Low energy adaptive clustering hierarchy with deterministic cluster-head selection, *4th IEEE international workshop on mobile and wireless communications network*, 2002, **18**, 368–372, doi: 10.1109/MWCN.2002.1045790.

[18] N. S. Patil, A. Parveen, *International Journal of Information Technology*, 2021, **20**, 1-10, doi: 10.1007/s41870-021-00791-y.

[19] M. Radhika, P. Sivakumar, *Wireless Networks*, 2021, **27**, 27-40, doi: 10.1007/s11276-020-02435-8.

[20] R. C. Shah, S. Roy, S. Jain, W. Brunette, *Ad-hoc Networks*, 2003, **1**, 215–233, doi: 10.1016/S1570-8705(03)00003-9.

[21] S. D. Kumar, H. Ahmad, A. Uniyal, Reliable ILP approach of Max-RWA problem for translucent optical network, *Thirteenth IEEE International Conference on Wireless and Optical Communications Networks (WOCN),* 2016, doi: 10.1109/WOCN.2016.7759879.

[22] A. O. Abu Salem, N. Shudifat, *Personal and Ubiquitous Computing*, 2019, **23**, 901-907, doi: 10.1007/s00779-019-01205-4.

[23] M. Saxena, A. Joshi, S. Dutta, K. C. Mishra, A. Giri, S. Neogy, *Wireless Personal Communications*, 2021, **23**, 1-14, doi: 10.1007/s11277-021-08140-9.

[24] P. Ullas, K. S. Shivaprakasha, *Journal of Telecommunications and Information Technology*, 2021, **12**, 123-155, doi: 10.26636/jtit.2021.147420.

[25] X. Li, M. Nayak, I. Stojmenovic, Geographic routing in wireless sensor and actuator networks, *Wireless Sensor and Actuator Networks: Algorithms and Protocols for Scalable Coordination and Data Communication*, 2010, **10**, 153-159, doi:10.1002/9780470570517.

[26] Y. Yuanyuan, M. Ma, Data gathering in wireless sensor networks with mobile collectors, *IEEE International Symposium on Parallel and Distributed Processing*, 2019, 1–9, doi: 10.1109/IPDPS.2008.4536269.

Author Information

Jainendra Singh *is Research Scholar in the Department of Electrical Engineering, Faculty of Engineering & Technology, Jamia Millia Islamia (A Central University), New Delhi, India. He received his B.Tech. Degree in Computer Science & Engineering from Uttar Pradesh Technical University, Lucknow (UP) and M.Tech. (CTA) Degree from Delhi Technological University Delhi. He has published various research papers in International Journals and Conferences in the areas of Wireless Sensor Networks.*

Zaheeruddin *is currently Professor in the Department of Electrical Engineering, Faculty of Engineering & Technology, Jamia Millia Islamia (A Central University), New Delhi, India since 2003. He was Head of the Department of Electrical Engineering from 2011 to 2014. He joined the Department as a Lecturer from the very beginning of its establishment in 1986. He received the B. Sc. Engg. Degree in Electrical and M. Sc. Engg. Degree in Electronics & Communication from Aligarh Muslim University (AMU), Aligarh (UP) in 1982 and 1988 respectively and a Ph. D. Degree in Computer Science and Technology from Jawaharlal Nehru University (JNU), New Delhi in 2002. Prof. Zaheeruddin is a Fellow of IETE (India) and Life Member of World Federation of Soft Computing (U.S.A.), The Institution of Engineers (IE, India), Computer Society of India (CSI), Indian Society for Technical Education (ISTE), The Indian Science Congress Association (ISCA), and Indian Chapter of The International Centre for Theoretical Physics (Indian Chapter of ICTP). Zaheeruddin has published eighty (80) research papers in International Journals and Conferences in the areas of Artificial Intelligence, IoT, Soft Computing, Noise Pollution, Wireless Sensor Networks, Optimization Techniques, and Smart Grid. He is Reviewer of many International journals of IEEE, Elsevier, Springer, and ACTA Press etc.*