Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 9494476, 15 pages
https://doi.org/10.1155/2022/9494476

Research Article

SDBMND: Secure Density-Based Unsupervised Learning Method with Malicious Node Detection to Improve the Network Lifespan in Densely Deployed WSN

Tripti Sharma,[1] **Amar Kumar Mohapatra**,[2] **and Geetam Tomar**[3]

[1]*Research Scholar Indira Gandhi Delhi Technical, University for Women, IT Department, Faculty Maharaja Surajmal Institute of Technology, IT Department, New Delhi, India*
[2]*Indira Gandhi Delhi Technical, University for Women, IT Department, New Delhi, India*
[3]*Rajkiya Engineering College, Sonbhadra, Uttar Pradesh, India*

Correspondence should be addressed to Tripti Sharma; tripti_sharma@msit.in

Received 16 February 2022; Revised 5 March 2022; Accepted 9 March 2022; Published 28 March 2022

Academic Editor: Kalidoss Rajakani

Random deployment, the absence of central authority, and the autonomous nature of the network make wireless sensor networks (WSNs) prone to security threats. Security, bandwidth, poor connectivity, intrusion, energy constraints, and other challenges are critical and could affect the performance of the WSN while considering the energy-efficient and secure routing protocols in WSNs. Security threats to WSNs are gradually being expanded. Thus, to improve the network's performance, detection of anomalies (malicious and suspicious nodes, redundant data, bad connections, etc.) is important. This paper is aimed at introducing the malicious node detection algorithm based on the DBSCAN algorithm, which is a density-based unsupervised learning method for enabling wireless sensor networks to be much more secure and reliable. The prime objective of this algorithm is to develop a routing algorithm capable of detecting malicious nodes and having a prolonged network lifespan and higher stability period. Clustering and classification are two well-known methods in the field of machine learning that can be successfully used in various domains. Density-based clustering is a popular and extensively used approach in various domains. The DBSCAN is the utmost popular and best-known density-based clustering algorithm and is capable of determining arbitrary-shaped clusters. This paper addresses the two anomalies in the WSN, namely, spatial redundancy and malicious node identification. In this article, an algorithm has been suggested to reduce redundant data transmission along with the identification of suspicious nodes to conserve energy and to avoid falsification of data through malicious nodes. The analysis of simulation results and comparison of other algorithms that are in the same class shows that the SDBMND performs significantly better than EAMMH, TEEN, IC-ACO, and LEACH in dense networks.

1. Introduction

Modern communication requires secure and high-quality data transmission. With the advancement of micro-electro-mechanical systems, cheap and tiny sensor nodes (SNs), the wireless sensor network has emerged as a widely used field in various applications such as medical science, industry, agriculture, and environmental home applications [1]. In most applications, the SNs have a vital role in decision-making; hence, the quality of data and quick response are the prime concerns in a WSN [2]. The presence of data that seems inconsistent with the rest of the data set is considered anomalous data. Many anomaly detection and prevention-based techniques have come into existence since the security threats to WSN are increasing day by day. The advancement of anomaly detection techniques is important in WSN as these networks are characterised by constrained resources like limited energy, limited memory, short communication range, poor connectivity, bandwidth, computation, and capability. Anomalous data also referred to as outliers are unusual measurements [3, 4].

Anomaly detection is a crucial issue in WSN that demands precise, effective, and timely data analysis to aid

critical decision-making in various applications. In WSNs, SNs are densely deployed; hence, there is a high likelihood of sensing and transmitting redundant information. The redundant data transmission affects the residual energy of the SNs. Thus, it impacts the network performance as energy is the prime constraint in WSN.

Redundancy means a supplement or replica of resources that leads to similar data obtained from different resources. In a wireless sensor network, redundancy has a negative as well as a positive impact. To have reliable and quality data, redundancy plays a vital role. On the other hand, duplicate data transmission and acquisition lead to speedy energy exhaustion [5]. A literature survey reveals that a lot of effort is required to minimise redundant data transmission to increase the lifespan of WSN. Various redundancies, like spatial redundancy, temporal redundancy, physical redundancy, and analytical redundancy, have been addressed in a wireless sensor network. In spatial redundancy, similar information has been acquired from a different resource. In the case of a densely deployed WSN, the SNs offer spatial redundancy, which delivers a huge amount of redundant data [6]. In this paper, we are considering the densely deployed WSN; hence, spatial redundancy has been addressed. In order to have an efficient routing protocol for a densely deployed network, it is important to deal with the problem of spatial redundancy because it takes up a lot of network resources and increases the amount of network overhead.

WSNs are susceptible to various threats. It is very important to identify and remove the compromised or malicious nodes to avoid having falsified information or data being introduced by the adversary through these SNs. A series of threats, including buffer overflow, big mouth attack, and DoS attack, may occur due to mischievous nodes in WSNs. Since WSN has constrained working conditions and is vulnerable to various security threats, the malicious nodes may impact the performance of the entire network. Hence, the identification of these malicious nodes in WSN has become a research hotspot [7, 8]. Existing routing algorithms such as LEACH, EEAMH, TEEN, and IC-ACO cannot handle malicious nodes or spatial redundancy.

In this paper, the three key issues: redundant data transmission, identification of malicious nodes, and improvement of network life span by selecting the optimal route, have been addressed as prime objectives, which are the prime concerns in the design of energy-efficient and secure routing algorithms in WSN. The descriptions of the key contributions are as follows:

(1) Efforts have been made to develop a routing algorithm capable of handling redundant data. The DBSCAN method is applied to divide the network area into low- and high-density cluster regions. The nodes in a high-density area are prone to sense the redundant information. In high-density regions, efforts have been made to handle redundant information. In low-density areas, critical information is handled

(2) The malicious nodes have been identified based on the temperature values. All the possible malicious nodes will not be able to send the sensed data in the next rounds. This will save energy and make the data more secure and accurate

(3) Hence, since these two issues add extra overhead in handling the redundancy and detection of malicious nodes, they affect the lifespan of the network. Hence, efforts have been made to increase the network lifespan by simulating the behaviour of ants by incorporating the ACO (ant colony optimization) algorithm to route information in various clusters from SNs to cluster heads (CHs) and from CHs to sink node or base station (BS)

The rest of the paper is organised as follows. Section 2 talks about existing techniques in depth. Section 3 gives an overview of the radio model used in the implementation, and Section 4 talks about the SDBMND in depth. In Section 5, the simulation results have been discussed and the comparison of the SDBMND has been done with the TEEN, EEAMH, IC-ACO, and LEACH algorithms. The performance has been assessed based on various parameters. Lastly, Section 6 concludes the research work and discusses future work.

2. Literature Review and Related Work

Due to low-quality sensors, signal interference, unsupervised nature, harsh environmental conditions, and other dominant factors, sensors grieve with the anomaly in receiving information.

A lot of energy is wasted in the processing of this anomalous and redundant information [9]. Thus, in a dense network, detecting anomalies and malicious nodes, identifying the appropriate CH, and selecting the best path can significantly improve the routing algorithm's performance.

Various hierarchical routing algorithms have been proposed in the literature. However, the most well-known and traditional routing method is LEACH, which is widely employed by academics in their research. LEACH [10] is a cluster-based hierarchical routing technique. CHs are used to deliver processed data to the BS, and clusters are generated based on the strength of the received signal. Only such CHs are used for transmission, which helps the network's nodes save energy. Data processing operations like aggregation and data fusion are done locally within the cluster.

The literature review reveals that the appropriate cluster selection using fuzzy logic can enhance the performance of the routing algorithm in WSN. Zadeh [11] introduces a natural extension of ordinary set theory and names it fuzzy set theory. The literature survey also demonstrates that, using fuzzy set theory, different classical problems can be turned into their fuzzy equivalents, allowing them to be employed more effectively in a variety of applications.

FMCHEL [12] is a homogeneous routing algorithm based on fuzzy logic. All the SNs are homogeneous in nature. This algorithm was developed for the scenario where

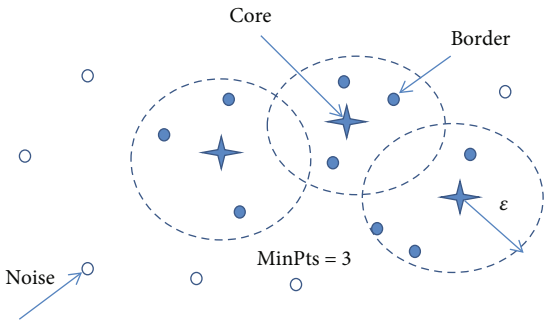

Figure 1: Illustration of DBSCAN points.

the base station is positioned far away from the network area. In this algorithm, in order to maximise the network lifespan, fuzzy logic-based cluster CH selection is performed. The FMCHEL is substantially more energy-efficient due to the concept of selecting a master cluster head among the selected CHs. The processed data can only be sent to BS by the master cluster head. The master-cluster-head node transfers the processed information to the BS. In CHEF [13], the selection of CHs is done in a distributed manner. The CHEF follows a fuzzy-based approach for the CH selection. While selecting the cluster heads, this algorithm also guarantees that no two CHs should lie within r distance. In CHEF, the node with the maximum energy and that is locally optimum is chosen as CH.

Bandyopadhyay and Coyle [14] proposed EEHC as a hierarchical clustering algorithm with K-steps. It is distributed in nature. This algorithm was made to solve the problems of one-hop random selection routing protocols like C-LEACH [15] and LEACH by adding multihop cluster formation to the process.

Fathi [16] applied the base DBSCAN algorithm for the formation of clusters, and cluster head selection is based on two parameters, namely, distance and energy. The simulation results demonstrate that as energy is saved in each round, the total network lifespan is improved in comparison to existing routing protocols.

In IC-ACO [17], a solution is proposed to implement and design a routing algorithm that has the ability to avoid random transmission. The deployment of SNs is random. The ACO approach has been used for selecting the optimal route for data transmission. The performance of the IC-ACO is compared, and simulation results show that the IC-ACO performs better in densely deployed networks and is more energy-efficient in dense networks. Hence, it can be successfully applied to densely deployed WSNs.

Abbasi and Younis [18]and Afsar and Tayarani-N [19] compared and reviewed various algorithms and protocols. Furthermore, some additional existing algorithms, like FLOC [20], MOCA [21], PEGASIS [22], HEED [23, 24], CCHs [25], EEDC [26], EEMC [27], CA-GA [28], DSCBA [29], LMANET [30], LCM [31], P-SEP [32], are among the latest work in the area of clustering.

TEEN [33] is a threshold-sensitive energy-efficient routing protocol. It is an improvement to the LEACH protocol. The TEEN protocol introduces the ideas of "soft threshold" (ST) and "hard threshold" (HT) to restrict the number of data transmissions.

EAMMH [34] is a multihop, multipath energy-aware algorithm. This algorithm works in two stages. A setup phase followed by a steady-state phase is used for cluster formation and data transmission. The user needs to provide the number of nodes as input. After the node deployment, the nodes execute the neighbor discovery algorithms to find the neighbor nodes. The selected CHs broadcast the advertisement message to all neighboring nodes, and after that, cluster formation takes place within the fixed boundary.

3. Review of Existing Methodologies Used for SDBMND

3.1. DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Clustering is an unsupervised learning technique that splits the data points into various particular clusters or groups in such a way that data points that belong to the same clusters will have similar properties and data points in other clusters will have other properties. There are many clustering algorithms based on different distance measures, like affinity propagation, spectral clustering, mean-shift, Gaussian mixture, and DBSCAN algorithms [35]. For a range of different distributions, the DBSCAN algorithm also yields more realistic results than k-means. The main concept of density-based clustering, also known as unsupervised learning methods, is to classify distinguished clusters in the data in such a manner that clusters are separated from each other by contiguous areas of high and low point density. DBSCAN is a density-based algorithm in which clusters of different sizes and shapes from a large amount of data can be identified, which contains outliers and noise. The following two parameters have been used in DBSCAN:

(1) MinPts: it is the minimum number of data points grouped together to be considered dense

(2) Eps (ε): the distance used to trace the points in the neighborhood of any point

These two parameters can be well explained with the help of two concepts: density connectivity and density reachability.

Reachability in terms of density finds a point to be reachable from another if it lies within an Eps distance from it.

Connectivity comprises transitivity-based chaining approach to find out whether points are positioned in a particular cluster.

In the DBSCAN algorithm, three types of points have been considered as seen in Figure 1 which are discussed below.

Core: it is a point that has at least n points within a distance x from itself.

Border: it is a point that has at least one core point at a distance x.

Noise: it is neither a core nor a border point. This point has fewer than x points within a distance of x from itself.

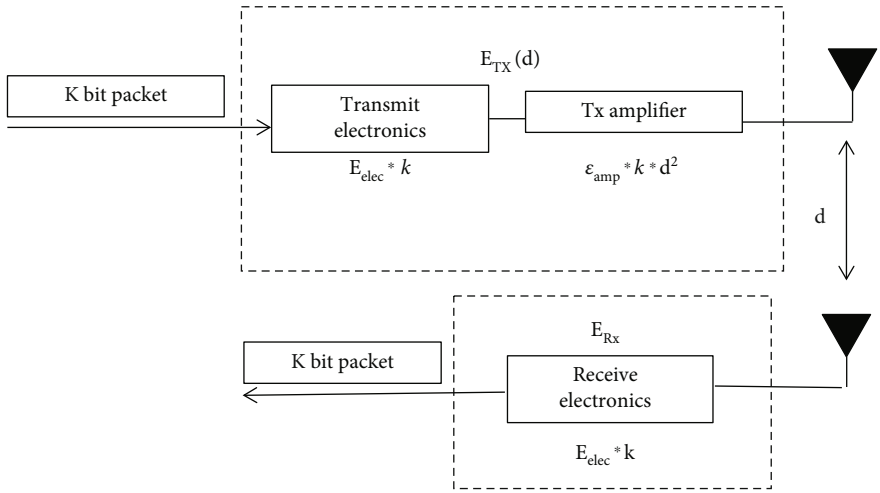

Figure 2: Energy dissipation diagram.

3.2. Ant Colony Optimization. It is a probabilistic method for answering computational problems that can be condensed to find the best route through graphs. These artificial ants are stimulated by the behaviour of real ants. Communication is based on the pheromone lying behaviour of biological ants. Artificial ants trace the optimal solution by traversing through the parameter space and by signifying all promising answers. Pheromones have been laid down by the real ants and direct each other to a final goal by traveling around within the environment.

The ACO algorithm was proposed in 1992 by Marco Dorigo [36]. The algorithm was designed with the objective to discover the optimal route in a graph entirely based on the behaviour of ants seeking a route between the food and their colony. This concept has now been expanded to unravel a broad range of numerical problems, which further causes various or numerous emerging fields portraying the numerous characteristics of the behaviour of ants.

In this algorithm, an ant is considered as a computational agent that tries to locate an optimal solution to a given optimization problem. To apply this approach to optimization issues, it must first be transformed into the problem of finding the shortest route [37, 38]. In the first step, every ant stochastically creates a partial solution. In the second step, the route identified by other ants is matched. Finally, on each edge, the pheromone levels get updated.

3.3. Energy Model Analysis. A simple energy model proposed by Heinzelman et al. is being cast off in this algorithm. The amplifier used for electronic receiving and transmission is displayed in Figure 2. Here, the power attenuation is based on the distance between the receiver and transmitter.

The energy dissipation for transmission is calculated as

$$\begin{aligned} E_{Tx}(k,d) &= E_{\text{elec}} * k + \epsilon_{fs} * k * d^2 \text{ if } d < d_0 \\ &= E_{\text{elec}} * k + \epsilon_{\text{mp}} * k * d^4 \text{ if } d \geq d_0. \end{aligned} \quad (1)$$

Table 1: Fuzzy sets and input variables.

S. no.	Input	Fuzzy sets		
1	SN's residual energy	Low	Average	High
2	SN's distance from BS	Near	Medium	Far

The energy dissipation on receiving a k-bit data packet is

$$E_{Rx}(k) = E_{\text{elec}} * k. \quad (2)$$

E_{elec} is the parameters for energy dissipation.
k is the packet size.
d denotes the distance between two nodes.
E_{fs} and E_{mp} are transmitter amplifier characteristics.

3.4. Combinatorics-Based Feature Analysis. This section discusses the feature analysis based on the principle of combinatorics that helps in determining the global dynamics of different parameters for malicious nodes in densely deployed WSN. The combinatorial basic can be written as

$$(x)^n = P_k^n = \frac{n!}{(n-k)!}. \quad (3)$$

The selection of parameter P can be done on the basis of k features from a set of n nodes and can be written as nk combination without repetition that can be written as

$$\binom{n}{k} = C_k^n = \frac{n!}{k!(n-k)!}. \quad (4)$$

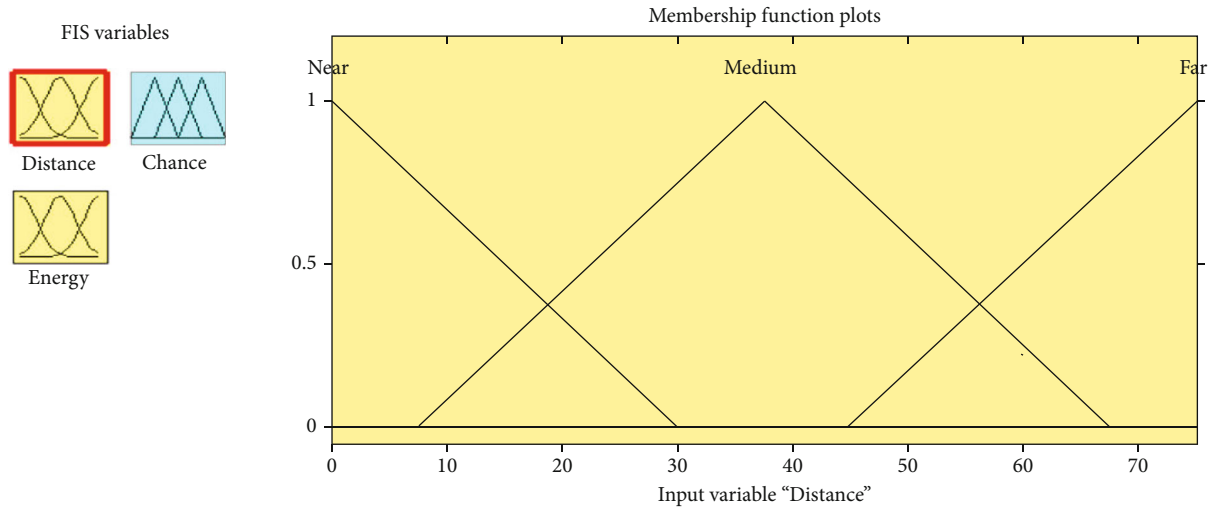

Figure 3: Membership function editor_distance.

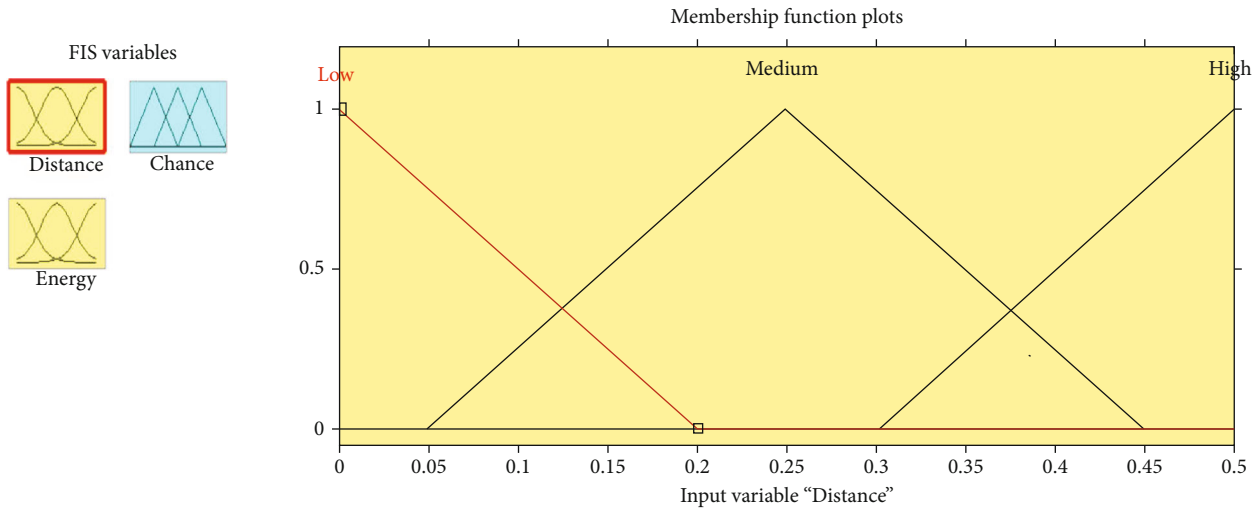

Figure 4: Membership function editor_residual energy.

For selecting k features from a set of n nodes, each node can be tested more than once. In this case, the different number of combinations can be

$$f_k^n = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}. \tag{5}$$

In equation (5), f_k^n can provide many integer solutions that can be written as

$$= p_1 + p_2 + p_3 + \cdots + p_n = k, \text{where } p_i \geq 0 \left(i \in \overline{1,n}\right). \tag{6}$$

To deduce the prediction, value k features so that the corresponding vector may differ in $\binom{n}{k}$ ways [39]. The subcomponent can differ in $2k$ ways that can be equal to 2 $n-k$ ways. With the help of permutation and combination with a repetition formula, after multiplying

$$\binom{n}{k} \bullet 2^k \bullet 2^{n-k} = \binom{n}{k} \bullet 2^n. \tag{7}$$

According to the inclusion-exclusion principle, which is also known as the sieve principle,

$$\left|\bigcup_{i=1}^{n} C_i\right| = \sum_{i=1}^{n} (-1)^{i-1} \sum_{\substack{I \subseteq 1,n \\ |I|=i}} \left|\bigwedge_{j \in I} C_j\right|. \tag{8}$$

Table 2: Inference rules.

S. no.	Distance	Energy	Chance
1	Near	Low	Low
2	Near	Average	High
3	Near	High	High
4	Medium	Low	Low
5	Medium	Average	High
6	Medium	High	High
7	Far	Low	Low
8	Far	Average	Low
9	Far	High	High

The above equation (8) can be restructured for two parameters (X) and Y, which can be generalised as

$$|X \cup Y| = |X| + |Y| - |X \cap Y|. \tag{9}$$

Equation (8) can now be extended as

$$\left|\bigcup_{i=1}^{n} C_i\right| = \sum_{i=1}^{n} (-1)^{i-1} \sum_{\substack{I \subseteq 1,n \\ |I|=i}} |C_I| = \sum_{i=1}^{n} (-1)^{i-1} \frac{n!}{i!} = \sum_{i=1} \frac{(-1)^{i-1}}{i!}. \tag{10}$$

And, from equations (8) and (10), it can be deduced as

$$|C_0| = |C| - \left|\bigcup_{i=1}^{n} C_i\right| = n! - n! \sum_{i=1}^{n} \frac{(-1)^{i-1}}{i!} = n! \sum_{i=0}^{n} \frac{(-1)^{i}}{i!}. \tag{11}$$

Lastly,

$$n! \sum_{i=0}^{n} \frac{(-1)^{i}}{i!} = \left[\frac{n!}{e}\right]. \tag{12}$$

The problems associated with the detection of certain diseases can be better identified with the help of combinatorics. This also helps in finding combinations of parameters in various features of certain components for densely deployed WSN.

4. Detailed Discussion for SDBMND

The proposed hierarchical unsupervised clustering-based routing algorithm is discussed in this section. The algorithm has four phases. The first phase includes the separation of densely deployed and sparsely deployed sensor network areas using the DBSCAN algorithm and the selection of CH nodes. The second phase includes the identification of malicious nodes and the removal of those malicious nodes from the network in order to avoid the introduction of falsified information within the WSN. In the third phase, sleep management is applied in order to avoid spatial redundancy. Sleep management has been incorporated to avoid redundant data transmission in a densely deployed network. The fourth step is the transmission phase, in which data is sent from node to CHs and from CHs to BS in an optimized manner based on ant behaviour using ACO algorithm.

4.1. The First Phase. In this phase, the setup of the network and the separation of densely and sparsely deployed sensor network areas have been completed. The network is randomly deployed, so some of the network areas are densely deployed and some of the network areas are sparsely deployed. Sleep management is applied to avoid redundant transmission and to save the energy of SNs in the transmission of similar data in densely deployed networks, but at the same time, we cannot overlook the SNs in sparsely deployed networks because they are critical and cover the information of those areas.

4.2. Separation of Sparse and Dense Network Regions Using the DBSCAN Algorithm and Selection of CHs Based on Fuzzy Logic. Like other clustering approaches, DBSCAN is a density-based unsupervised learning method to discover data proximity. The Euclidean distance has been employed to calculate the distance in this approach. Clustering is done using two parameters: Eps and MinPts. Eps is the neighborhood's maximum radius, and MinPts is the cluster's minimal number of SNs in a given area [40]. This algorithm determines the clusters as low-point and high-point density zones. It has the detection capability of an arbitrarily shaped cluster set along with the ability to locate the noisy data. It is difficult to estimate the proper value of Eps in the case of DBSCAN. If we select a large value of Eps, then finding the noisy data point is difficult and the number of clusters formed is also lower. If we select the smaller value for Eps, a large number of clusters will be formed, and many points will be considered noisy. The DBSCAN algorithm is used to separate the low-density and high-density areas in this phase.

4.3. Fuzzy Logic-Based CH Selection. In this phase, the CHs are chosen using a fuzzy logic-based approach. For the identification of optimal CHs, the distance between SNs and the BS and residual energy was evaluated as input factors [41]. The network plots in a range of $100 * 100$, where the value of the distance could be anywhere from 0 to 75. The distance could be represented as three fuzzy sets: far, medium, and near. Table 1 represents the fuzzy sets for energy and distance [42]. The membership function for distance and residual energy is represented in Figures 3 and 4. The proposed approach is homogeneous, so all the SNs have similar initial energy. Hence, the value of residual energy could lie anywhere between 0 and 0.5, as shown in Figure 4. The Mamdani rule-based model is used by the SNs for decision-making, and existing fuzzy if-then rules have been used in the selection of the neighboring node as chances [43]. The rule base and its representation are given below in Table 2.

When the data is reformed to the membership functions and additional functions are reassigned to the fuzzy inference, the fuzzy rule base [44, 45] copes with these linguistic factors to provide the output.

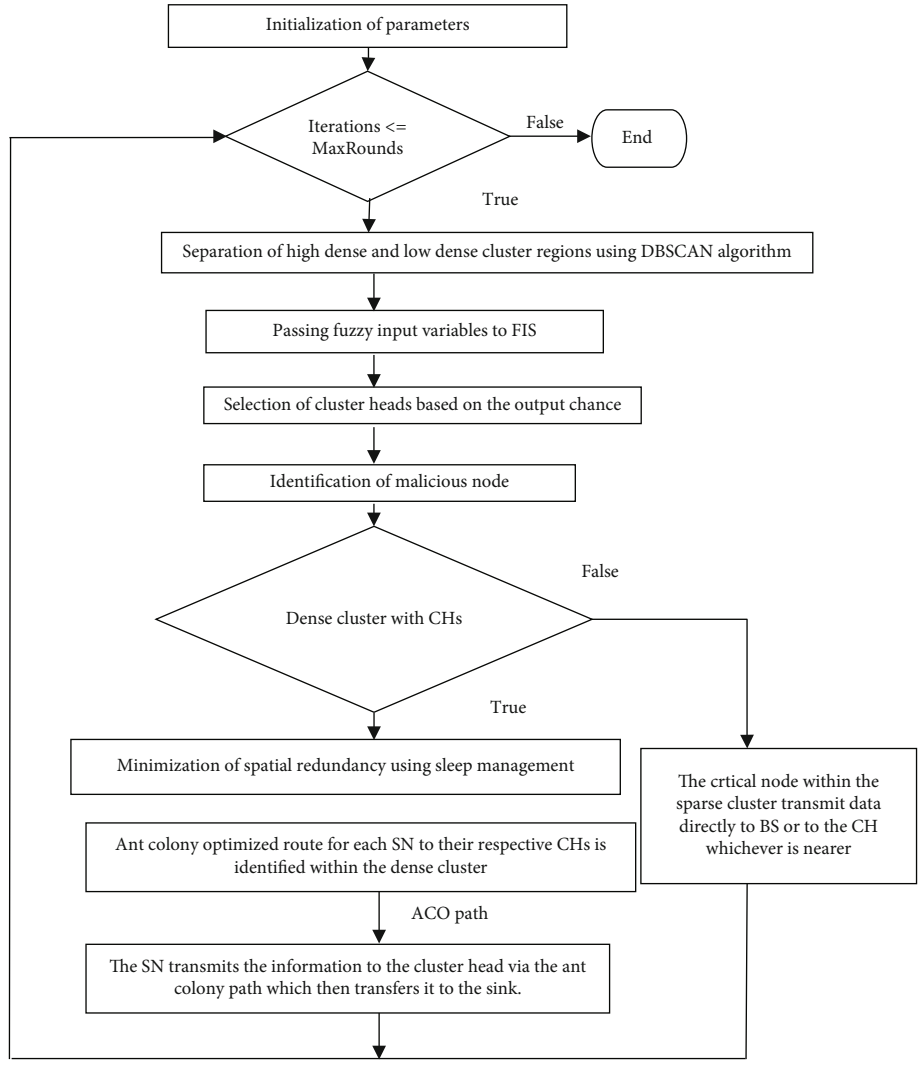

Figure 5: Flowchart for the proposed approach.

4.4. Second Phase (the Identification of Malicious Nodes). In this phase, every SN senses the temperature value and has an initial weight value. Initially, the value of weight is set to 1 for all the SNs. Every node transmits the weight value and temperature value to the CHs [46]. To identify the malicious node, the optimum temperature value must be determined. It could be calculated using

$$\text{Topt} = \frac{\left(\sum_{i=1}^{n} wi * \text{Temp}i\right)}{\sum_{i=1}^{n} wi}. \tag{13}$$

Here, Topt is the weighted average temperature, i.e., initially all the nodes have a weight value of 1.

There could be some events, and because of those events, the temperature of SNs could be increased or decreased, so there would be temperature variation. To find the range for trusted optimal temperature (Topt + var) < Topt_trusted > (Topt-var), some variation is allowed. If any node transmits the temperature value within this limit, then that node is considered a normal node. Otherwise, it would be considered a probable malicious node and the weight of those SNs would get updated with some fine as

$$wi = wi - \beta. \tag{14}$$

Otherwise, values will remain the same. β is fine or penalty.

Table 3: Parameter values.

Simulation parameters	Values
The BS's (XY) position	$(50, 50)$
Maximum rounds	3000
$E_{\text{TX}} = E_{\text{RX}}$	$50 * 0.000000001$ joules
E_{mp}	$0.0013 * 0.000000000001$ joules
E_{fs}	$10 * 0.000000000001$ joules
Initial energy	0.5 joules
Data aggregation energy (EDA)	$5 * 0.000000001$ joules

Table 4: Round values at which the first node dies for LEACH, TEEN, EAMMH, IC-ACO, and SDBMND.

FND_round values	LEACH	TEEN	EAMMH	IC-ACO	SDBMND
FND_round value with 100 SNs	436	452	370	930	1570
FND_round value with 200 SNs	222	378	203	948	1620
FND_round value with 300 SNs	133	590	145	962	1690

Table 5: Improvement in stability period in SDBMND over LEACH, TEEN, EAMMH, and IC-ACO.

Description	An improvement over LEACH in %	An improvement over TEEN in %	An improvement over EAMMH in %	An improvement over IC-ACO in %
With SNs: 100	260	247	324	69
With SNs: 200	629	328	698	71
With SNs: 100	1170	186	1065	76

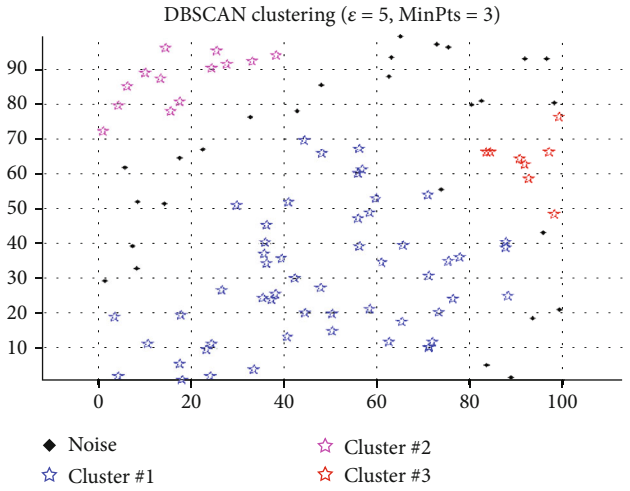

Figure 6: Formation of the cluster with 100 SNs.

This value keeps on being updated, and if this value goes below some threshold value (α), the node is considered a malicious node. The node will not participate further and be declared as malicious node.

4.5. Third Phase (Sleep Management to Avoid Redundant Data Transmission). The nodes are partitioned into low- and high-density regions using the DBSCAN algorithm. To avoid the transmission of false or wrong information, the malicious nodes identified in phase two will not be permitted to transmit the information. Since the nodes in the highly dense cluster tend to transmit the redundant information to the CHs, only 5% of the nodes with maximum energy are permitted to transfer the information, and the rest of the SNs will be in sleep mode. Only those 5% of nodes will sense the information, process the information, and be allowed to transmit that information to CHs. Hence, there are two advantages to using sleep management. First, the redundant information will not be transmitted; thus, the energy of nodes in sensing, processing, and transmission will be saved [47]. The same applies to CHs as well. The energy

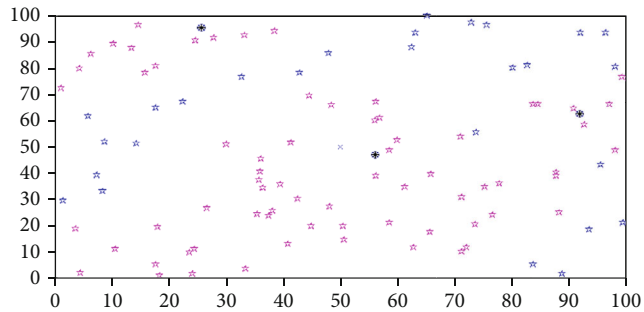

Figure 7: Selection of CH with 100 SNs.

of CHs is saved in the processing of redundant information received from all the nodes. The second advantage is that, since the energy of SNs is saved, they will remain alive for a longer period of time. This means the network will last longer as a whole.

4.6. Fourth Phase (the Transmission Phase Based on ACO). After the partitioning of the network into low- and high-density regions, the identification of malicious nodes, and the minimisation of redundant data transmission, the routing of data to BS takes place in the fourth phase. Since, in high-density areas, we have CHs, so they will transfer the processed information to the BS. In low-density areas, all the nodes are critical since they are sparse and they do not tend to sense redundant information. Hence, the information sensed by these nodes is critical and needs to be sensed, processed, and transmitted directly to the BS or CHs, whichever is nearer. In this phase, we applied the ACO algorithm to discover the best route for the transmission of data from SNs to CHs and CHs to BS in high-density regions. In this phase, the critical SNs also transmit the data to the BS in low-density regions.

4.7. The Emerging Solution for ACO-Based Routing. Information transmission within the network took place in this phase. The ACO algorithm is used to determine the best

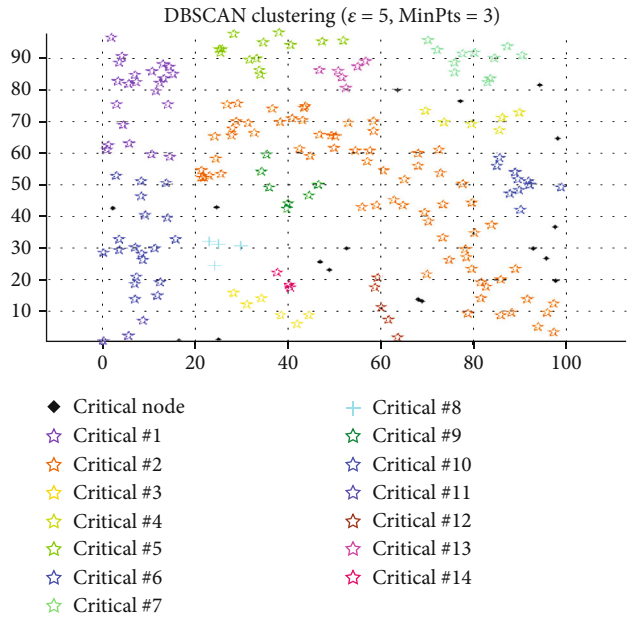

Figure 8: Formation of the cluster with 200 SNs.

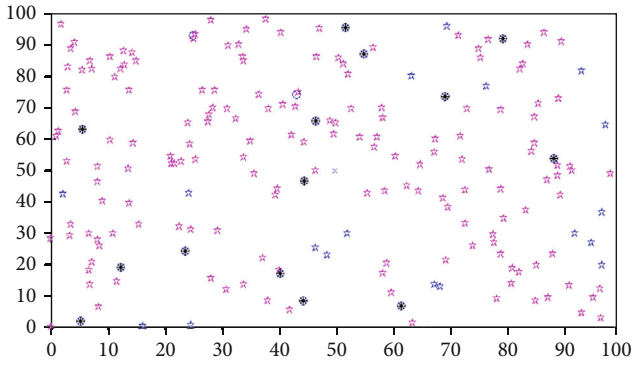

Figure 9: Selection of CHs with 200 SNs.

path between the SNs and CHs. The following stages are used to determine the best path between SNs and CHs:

(A) At each node, a forward ant is introduced

(B) To get to specific CHs, all ants practice the intermediate SNs with a specific goal

(C) The ants take a probabilistic approach to determine the next node to be navigated. The heuristic and pheromone data are used to form this probabilistic attitude. The probability is expressed as

$$P_{ij} = \frac{(\tau_{ij})^{\alpha 1}(\eta_j)^{\beta 1}}{\sum_{j \epsilon N}(\tau_{ij})^{\alpha 1}(\eta_j)^{\beta 1}}, \quad (15)$$

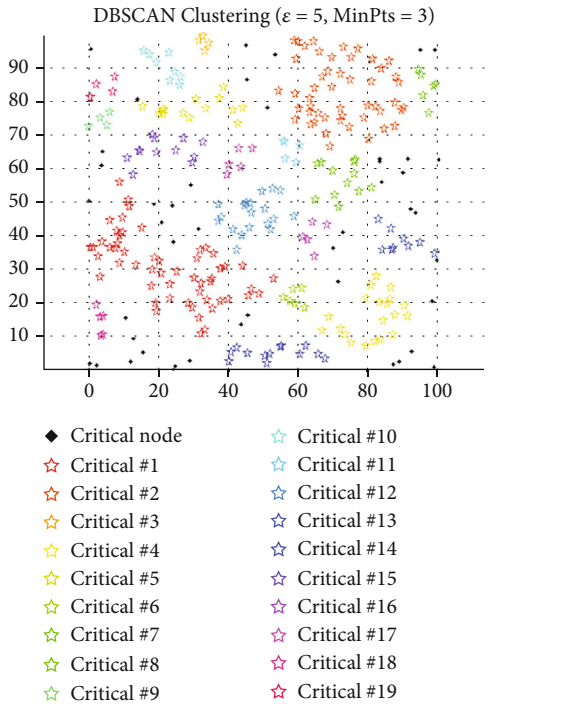

Figure 10: Formation of the cluster with 300 SNs.

where d_{ij} is the distance between the CH and SNi and τ_{ij} represents the pheromone information; it is calculated as

$$\tau_{ij} = \frac{1}{d_{ij}}. \quad (16)$$

η_j signifies the heuristic information, it denotes the node's energy, and it is figured as

$$\eta_j = \frac{E_0 - E_{\text{residual}}}{\sum_{k \epsilon N} E_k}, \quad (17)$$

where E_0 is the initial energy and E_{residual} is the residual energy. The α_1 and β_1 parameters aid in amending the relative weight of the pheromone trail and heuristic individually.

(D) An SN with maximum probability is selected as a succeeding hop to transfer the information to its allied CHs

4.8. Steps of the SDBMND

Step 1. Create a 100 ∗ 100 network area with n (100,200,300) nodes.

Step 2. As the nodes are randomly deployed, some of the nodes are densely deployed (highly dense) and some are sparsely deployed (low density). The highly dense areas and low-dense areas are separated using the DBSCAN algorithm.

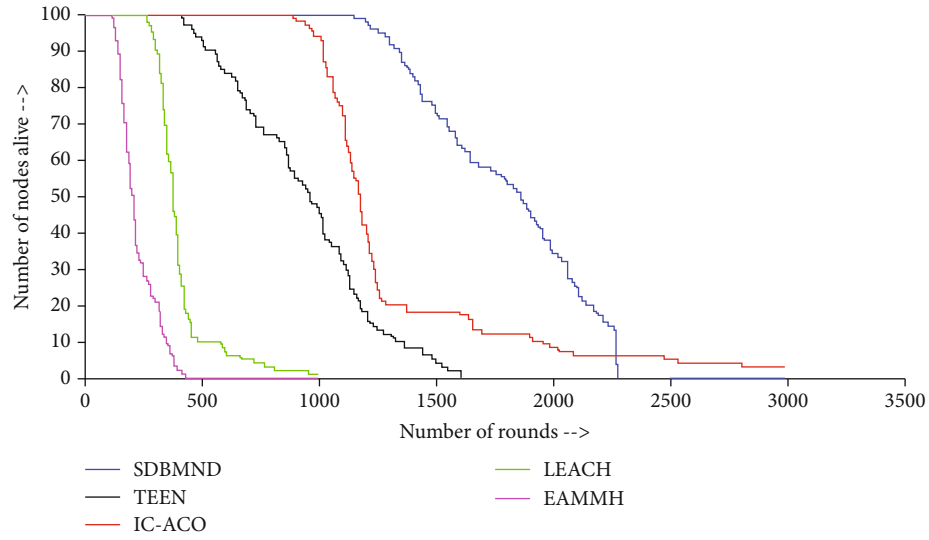

Figure 11: Number of SNs alive at different rounds (SNs: 100).

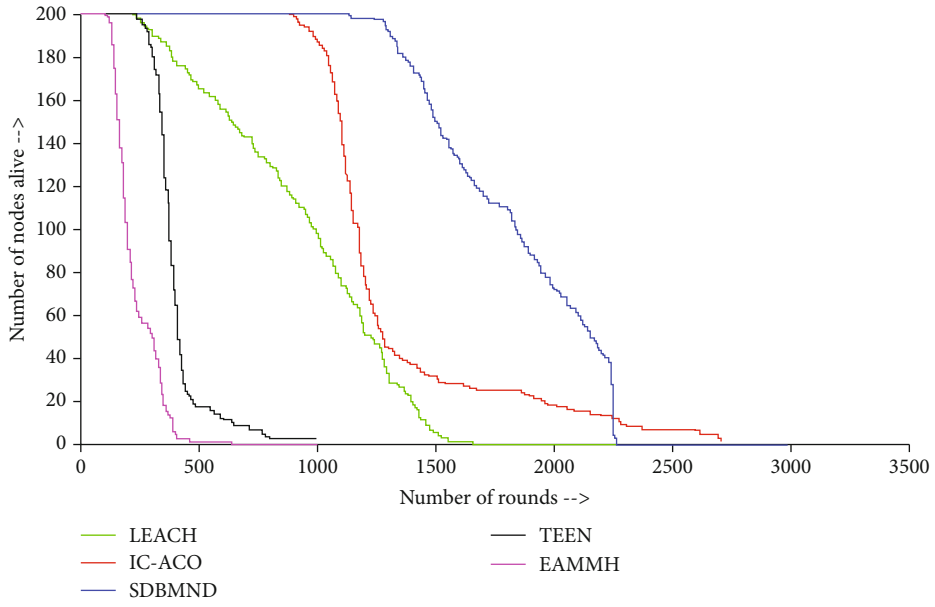

Figure 12: Number of SNs alive at different rounds (SNs: 200).

Step 3. In a densely populated area, there is a proclivity to detect similar information (redundant information, one of the anomalies). In low-density areas, nodes are sparse and are considered critical nodes. With the help of two fuzzy input parameters, the CHs have been identified in a high-density region. All the nodes in low-density regions are classified as "critical nodes."

Step 4. Every node senses the temperature value from its surroundings and has an initial weight value. Initially, the weight is 1 for all. Every node transmits the weight value and temperature value to the CH. Using equation (13), the optimum value of temperature has to be calculated to find the malicious node. There could be some events, and because of those events, the temperature of nodes could be increased or decreased, so there would be temperature variation. To find the range for trusted optimal temperature (Topt+var) (Topt-var), some variation is allowed. If any node transmits the temperature value within this limit, then that node is considered a normal node. Otherwise, it would be considered a probable malicious node and the weight of the nodes would be updated with some fine using equation (14). Otherwise, values will remain the same.

This value keeps on being updated, and if this value goes below some threshold value (α), the node is considered a malicious node. The node will not participate further and be declared a malicious node.

Step 5. As a result of being close together and being in a dense cluster, some of the nodes tend to sense redundant

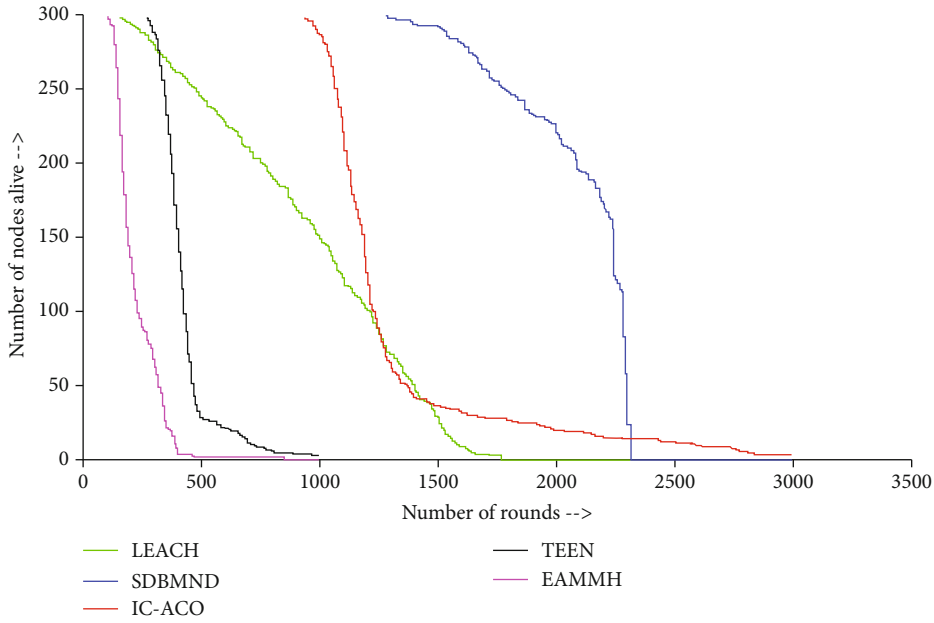

Figure 13: Number of SNs alive at different rounds (SNs: 300).

data. The nodes in the dense cluster follow the sleep management.

Step 6. Routing of data through ACO algorithm within the dense cluster has been done.

Step 7. The data will be sent straight from the critical node to the BS or to the nearest CH, whichever is closer to the node.

Figure 5 below shows the flowchart for the proposed algorithm.

5. Simulation Results and Analysis

The 100, 200, and 300 SNs are randomly deployed in the 100 ∗ 100 simulation environment. The performance of the SDBMND is compared with four existing protocols, namely, EAMMH, TEEN, IC-ACO, and LEACH, in a dense environment. Based on two parameters, namely, the stability period and number of packets transmitted by SNs, the performance of EAMMH, LEACH, IC-ACO, TEEN, and the SDBMND is compared. Table 3 lists the simulation parameters.

Two parameters are used in this performance analysis: the stable region and the number of data packets that are sent. Both the parameters are described below.

(1) The stable region is the region within which all the SNs are alive

(2) The number of packets transmitted by the SNs

Simulation results reveal that in comparison to EAMMH, TEEN, LEACH, and IC-ACO, the proposed approach has an improvement in the stability period as well as in the overall network lifetime of the networks, and it is also found to be superior in terms of energy utilisation when we compared it with the TEEN, EAMMH, IC-ACO, and LEACH. Table 4 indicates the round value until all the nodes are alive in the SDBMND, IC-ACO, TEEN, EAMMH, and LEACH with 100, 200, and 300 SNs. Table 4 shows the value of the round at which the first node dies for LEACH, TEEN, EAMMH, IC-ACO, and SDBMND, and it could be seen that SDBMND has the highest value of all in all the cases, i.e., 100 SNs, 200 SNs, and 300 SNs. A stability period could be calculated from the region within which all the SNs are alive. It could be seen clearly that the SDBMND has the highest stability period among all. Table 5 shows the improvement in stability period in SDBMND over LEACH, TEEN, EAMMH, and IC-ACO. The IC-ACO is specifically designed for dense network and it could be seen that SDBMND has improvement of 69%, 71%, and 76% over IC-ACO algorithm in case of 100, 200, and 300 SNs.

Figures 6 and 7 shows the cluster formation and the selection of CHs with 100 SNs. Those nodes which are not part of any cluster are considered critical nodes. Figures 8 and 9 show the cluster formation and selection of CHs with 200 SNs. Figure 10 shows the formation of the cluster when the number of nodes is 300. Hence, on increasing SNs, the SDBMND performs much better when the network is dense.

From Figure 11, all the SNs alive at various rounds can be seen, which specifies the network lifetime when 100 nodes are installed. It is clear that the suggested algorithm's performance is much better as compared to EAMMH, TEEN, IC-ACO, and LEACH. In LEACH, all the nodes are alive till 436 rounds, in IC-ACO till 930 rounds, in TEEN till 396 rounds, in EAMMH till 370 rounds, and till 1570 in the SDBMND, which illustrates the substantial progress in the stable region.

Figure 12 portrays the total alive SNs at different rounds, which specifies the network lifetime when the network is much dense, as 200 nodes are installed. In dense environments, the SDBMND outperforms the TEEN, EAMMH, IC-ACO, and LEACH, as shown in Figure 12. In the LEACH

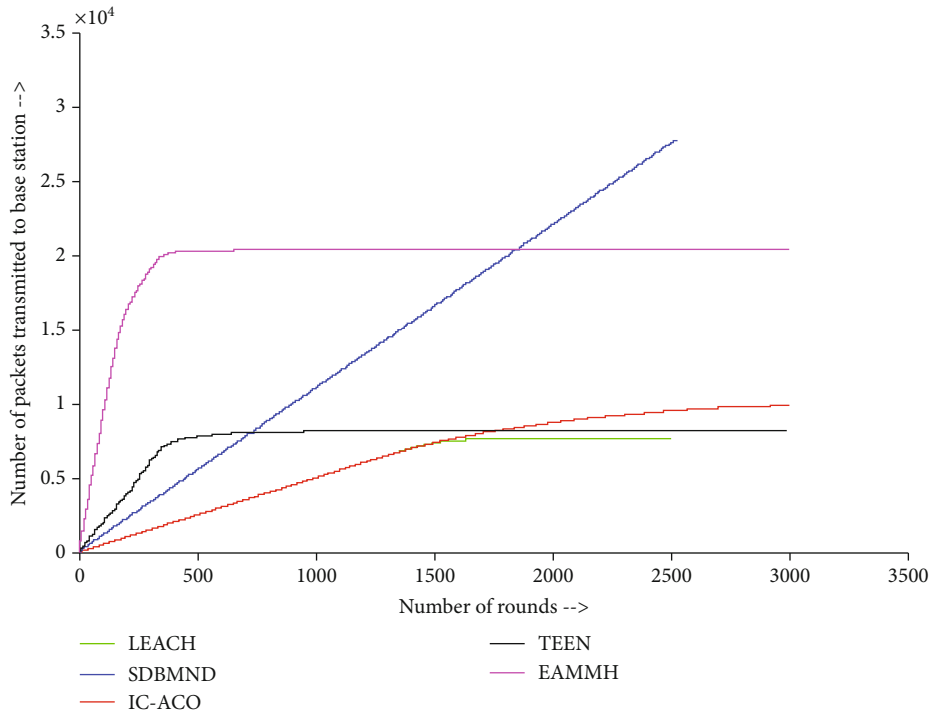

Figure 14: The total amount of packets that BS has received (SNs: 100).

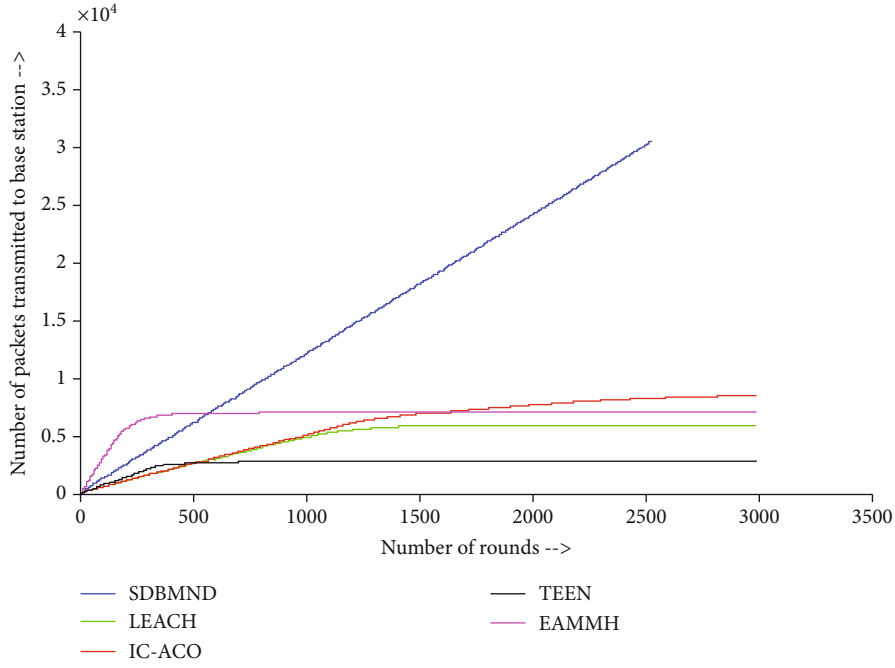

Figure 15: The total amount of packets that BS has received (SNs: 200).

protocol, all nodes stay alive until 222 rounds, 948 rounds in IC-ACO, 378 rounds in TEEN, 278 rounds in EAMMH, and 1620 rounds in the suggested algorithm. This shows that the LEACH protocol is less effective in dense networks, but the IC-ACO algorithm significantly improves its performance. However, SDBMND performed better than IC-ACO.

Figure 13 portrays the total alive SNs at different rounds, which specifies the network lifetime when the network is much dense, as 300 nodes are installed. Figure 13 illustrates that the SDBMND performs much better than IC-ACO and LEACH in dense environments. In the LEACH protocol, all nodes stay alive till 133 rounds, till 962 rounds in IC-ACO,

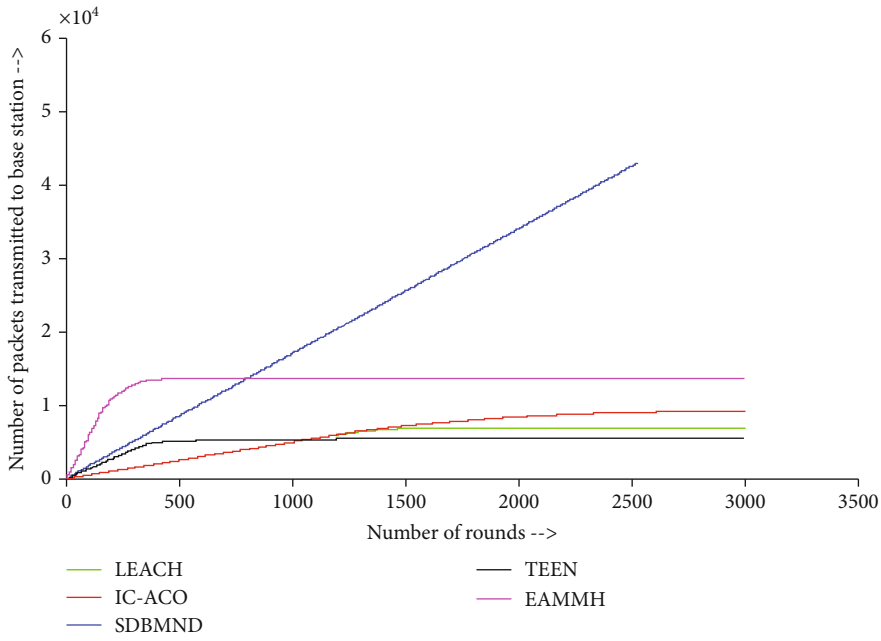

Figure 16: Total number of packets received by the BS (SNs: 300).

till 590 rounds in TEEN, till 120 rounds in EAMMH, and till 1690 in the suggested algorithm. The LEACH protocol's performance declined in a dense network, but the performance of the IC-ACO algorithm was significantly improved, whereas the SDBMND algorithm performed better than IC-ACO.

Figure 14 shows the data packets transmitted to the BS with 100 wireless SNs in the network. It could be seen that even though an effort has been made to avoid the transmission of redundant information, the total data packets received at the base station have increased. The SDBMND has a higher network lifetime in comparison to the LEACH, TEEN, EAMMH, and IC-ACO protocols.

Figure 15 shows the data packets acquired by the BS when 200 SNs are installed in the WSN. The figure shows that the SDBMND algorithm transmits more data packets to the BS than the existing TEEN, EAMMH, IC-ACO, and LEACH algorithms in a dense network.

Figure 16 displays the data packets acquired by the BS when 300 SNs are installed in the WSN. From the figure, it is evident that in the SDBMND the total data packets transmitted to the BS are higher compared to the existing TEEN, EAMMH, IC-ACO, and LEACH algorithms in a dense network.

Simulation results have the following conclusions.

SDBMND is much more stable than the TEEN, EAMMH, IC-ACO, and LEACH protocols when used in a densely deployed WSN.

The number of data packets transmitted has increased significantly for a similar network configuration.

6. Conclusion

Identification of malicious nodes, finding the optimal route, cluster formation, and processing of redundant information are some of the critical key problems in WSN. To extend the network's life span in densely deployed WSNs, the proposed study intends to save energy by avoiding redundant data transmission and identifying and deleting malicious nodes. When SNs are positioned in close proximity, there is a high probability of redundant data transmission. It is perceived that the SNs are typically positioned in close proximity in a dense network and are used to transfer redundant information; hence, energy is wasted in processing that redundant information. This approach is excellent for usage in a dense network since it avoids duplicate data transmission by employing the DBSCAN algorithm to separate low- and high-density regions, as well as proper CH selection using fuzzy logic, malicious node identification, and sleep cycle management. In the SDBMND, the malicious nodes have been identified to make this algorithm perform much better so that we can avoid the malicious information being processed by the BS and more accurate information can be received by the base station. The algorithm is designed to implement a secure routing algorithm that is competent in terms of stability period and prolonged network lifespan in a densely deployed network as compared to existing well-known algorithms like LEACH, TEEN, EEMAH, and IC-ACO, and simulation results show that the performance of the SDBMND is significantly better than the existing protocols, which makes it suitable to be used in a dense network. The IC-ACO algorithm is specifically designed for dense wireless sensor networks, but SDBMND shows 69%, 71%, and 76% improvements over IC-ACO with 100, 200, and 300 SNs, respectively. In the proposed work, all the SNs are homogeneous, but in the future, heterogeneous nodes could be considered. As a future scope, the size of the network could also be expanded to access the scalability of the proposed algorithm. In the proposed work, the temperature has been taken as a parameter to check the

malicious node. However, in the future, more parameters could be added to find out the probable malicious node.

Data Availability

The code has been implemented using MATLAB. The MATLAB code for the proposed work is available.

Conflicts of Interest

There are no conflicts of interest/competing interests among any author.

References

[1] C. C. Shen, C. Srisathapornphat, and C. Jaikaeo, "Sensor information networking architecture and applications," *IEEE Personal Communications*, vol. 8, no. 4, pp. 52–59, 2001.

[2] F. Fanian and M. K. Rafsanjani, "Cluster-based routing protocols in wireless sensor networks: a survey based on methodology," *Journal of Network and Computer Applications*, vol. 142, pp. 111–142, 2019.

[3] Y. Gao, K. Wu, and F. Li, "Analysis on the redundancy of wireless sensor networks," in *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications*, pp. 108–114, San Diego CA USA, 2003.

[4] H. S. Emadi and S. M. Mazinani, "A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks," *Wireless Personal Communications*, vol. 98, no. 2, pp. 2025–2035, 2018.

[5] N. Verma and D. Singh, "Data redundancy implications in wireless sensor networks," *Procedia Computer Science*, vol. 132, pp. 1210–1217, 2018.

[6] S. Kumar and V. K. Chaurasiya, "A strategy for elimination of data redundancy in the internet of things (IoT) based wireless sensor network (wsn)," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1650–1657, 2018.

[7] F. Zawaideh and M. Salamah, "An efficient weighted trust-based malicious node detection scheme for wireless sensor networks," *International Journal of Communication Systems*, vol. 32, no. 3, article e38788, 2019.

[8] F. Zawaideh, M. Salamah, and H. Al-Bahadili, "A fair trust-based malicious node detection and isolation scheme for WSNs," in *2nd International Conference on the Applications of Information Technology in Developing Renewable Energy Processes & Systems (IT-DREPS 2017)*, Amman, Jordan, 2017.

[9] J. Han, E. Lee, H. Cho, Y. Yoon, H. Lee, and W. Rhee, "Improving the energy-saving process with high-resolution data: a case study in a university building," *Sensors*, vol. 18, no. 5, 2018.

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

[11] L. A. Zadeh, "On fuzzy algorithms," in *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, World Scientific Publishing, 1996.

[12] T. Sharma and B. Kumar, "F-MCHEL: fuzzy based master cluster head election leach protocol in wireless sensor network," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 10, pp. 8–13, 2012.

[13] J. M. Kim, S. H. Park, Y. J. Han, and T. M. Chung, "CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks," in *10th International Conference on Advanced Communication Technology*, pp. 654–659, Gangwon, Korea (South), 2008.

[14] S. Bandyopadhyay and E. J. Coyle, "An energy-efficient hierarchical clustering algorithm for wireless sensor networks," in *Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEE IEEE INFOCOM 2003)*, San Francisco, CA, USA, 2003.

[15] R. Mehta, A. Pandey, and P. Kapadia, "Reforming clusters using C-LEACH in wireless sensor networks," in *2012 International Conference on Computer Communication and Informatics*, pp. 1–4, Coimbatore, India, 2012.

[16] M. Fathi and M. Nazari, "An energy-efficient density-based clustering approach for wireless sensor networks," *IJMEC*, vol. 7, no. 24, 2017.

[17] J. Y. Kim, T. Sharma, B. Kumar, G. S. Tomar, K. Berry, and W. H. Lee, "Intercluster ant colony optimization algorithm for wireless sensor network in dense environment," *International Journal of Distributed Sensor Networks*, vol. 10, no. 4, Article ID 457402, 2014.

[18] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826–2841, 2007.

[19] M. M. Afsar and M. H. Tayarani-N, "Clustering in sensor networks: a literature survey," *Journal of Network and Computer Applications*, vol. 46, pp. 198–226, 2014.

[20] M. Demirbas, A. Arora, V. Mittal, and V. Kulathumani, "Design and analysis of a fast local clustering service for wireless sensor networks," in *First International Conference on Broadband Networks*, pp. 700–709, San Jose, CA, USA, 2004.

[21] A. Youssef, M. Younis, M. Youssef, and A. Agrawala, "Wsn16-5: distributed formation of overlapping multi-hop clusters in wireless sensor networks," in *IEEE Globecom 2006*, pp. 1–6, San Francisco, CA, USA, 2006.

[22] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings, IEEE Aerospace Conference*, pp. 3–3, Big Sky, MT, USA, 2002.

[23] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.

[24] O. F. Younis, "Distributed clustering in ad—hoc sensor networks a hybrid energy-efficient approach," in *Proc 13th Joint Conference IEEE Computer and Communications Societies*, pp. 660–670, Hong Kong, China, 2004.

[25] H. Soleimani, S. Tomasin, T. Alizadeh, and M. Shojafar, "Cluster-head based feedback for simplified time reversal prefiltering in ultra-wideband systems," *Physical Communication*, vol. 25, pp. 100–109, 2017.

[26] M. M. Afsar and M. H. Tayarani-N, "A novel energy-efficient and distance-based clustering approach for wireless sensor networks," in *Soft Computing in Industrial Applications*, pp. 177–186, Springer, 2014.

[27] Y. Jin, L. K. Wang, Y. Kim, and X. Yang, "EEMC: an energy-efficient multi-level clustering algorithm for large-scale wireless sensor networks," *Computer Networks*, vol. 52, no. 3, pp. 542–562, 2008.

[28] M. Ahmadi, M. Shojafar, A. Khademzadeh, K. Badie, and R. Tavoli, "A hybrid algorithm for preserving energy and delay

routing in mobile ad-hoc networks," *Wireless Personal Communications*, vol. 85, no. 4, pp. 2485–2505, 2015.

[29] Y. Liao, H. Qi, and W. Li, "Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1498–1506, 2012.

[30] S. H. H. Nazhad, M. Shojafar, S. Shamshirband, and M. Conti, "An efficient routing protocol for the QoS support of large-scale MANETs," *International Journal of Communication Systems*, vol. 31, no. 1, article e3384, 2018.

[31] S. S. Wang and Z. P. Chen, "LCM: a link-aware clustering mechanism for energy-efficient routing in wireless sensor networks," *IEEE Sensors Journal*, vol. 13, no. 2, pp. 728–736, 2012.

[32] P. G. V. Naranjo, M. Shojafar, H. Mostafaei, Z. Pooranian, and E. Baccarelli, "P-SEP: a prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks," *The Journal of Supercomputing*, vol. 73, no. 2, pp. 733–755, 2017.

[33] A. Manjeshwar and D. P. Agrawal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," *ipdps*, vol. 1, no. 2001, p. 189, 2001.

[34] M. R. Mundada, V. CyrilRaj, and T. Bhuvaneswari, "Energy-aware multi-hop multi-path hierarchical (EAMMH) routing protocol for wireless sensor networks," *European Journal of Scientific Research*, vol. 88, no. 4, pp. 520–530, 2012.

[35] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *kdd*, vol. 96, no. 34, pp. 226–231, 1996.

[36] S. Okdem and D. Karaboga, "Routing in wireless sensor networks using an ant colony optimization (ACO) router chip," *Sensors*, vol. 9, no. 2, pp. 909–921, 2009.

[37] A. M. Zungeru, L. M. Ang, and K. P. Seng, "Classical and swarm intelligence based routing protocols for wireless sensor networks: a survey and comparison," *Journal of Network and Computer Applications*, vol. 35, no. 5, pp. 1508–1536, 2012.

[38] N. Moussa, E. Nurellari, and A. E. B. El Alaoui, "A novel energy-efficient and reliable ACO-based routing protocol for WSN-enabled forest fires detection," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2022.

[39] S. Ross, "A first course in probability," *Upper Saddle River*, vol. 6, 2009.

[40] D. Pan and L. Zhao, "Uncertain data cluster based on DBSCAN," in *2011 International Conference on Multimedia Technology*, pp. 3781–3784, Hangzhou, China, 2011.

[41] J. Anno, L. Barolli, A. Durresi, F. Xhafa, and A. Koyama, "A cluster head decision system for sensor networks using fuzzy logic and number of neighbor nodes," in *First IEEE International Conference on Ubi-Media Computing*, pp. 50–56, Lanzhou, China, 2008.

[42] S. Y. Chiang and J. L. Wang, "Routing analysis using fuzzy logic systems in wireless sensor networks," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 966–973, Zagreb, Croatia, 2008.

[43] J. S. Lee and W. L. Cheng, "Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication," *IEEE Sensors Journal*, vol. 12, no. 9, pp. 2891–2897, 2012.

[44] P. S. Mehra, M. N. Doja, and B. Alam, "Fuzzy based enhanced cluster head selection (FBECS) for WSN," *Journal of King Saud University-Science*, vol. 32, no. 1, pp. 390–401, 2020.

[45] T. Sharma, A. Mohapatra, and G. Tomar, "Fuzzy-based DBSCAN algorithm to elect master cluster head and enhance the network lifetime and avoid redundancy in wireless sensor network," in *International Conference on Innovative Computing and Communications*, pp. 1031–1042, New Delhi, 2020.

[46] I. M. Atakli, H. Hu, Y. Chen, W. S. Ku, and Z. Su, "Malicious node detection in wireless sensor networks using weighted trust evaluation," in *Proceedings of the 2008 Spring simulation multiconference*, pp. 836–843, Ottawa, Canada, 2008.

[47] H. M. Jawad, R. Nordin, S. K. Gharghan, A. M. Jawad, M. Ismail, and M. J. Abu-AlShaeer, "Power reduction with sleep/wake on redundant data (SWORD) in a wireless sensor network for energy-efficient precision agriculture," *Sensors*, vol. 18, no. 10, p. 3450, 2018.