International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

Energy Aware Fuzzy Logic Secure Data Aggregation (EA-FSDA) technique for Wireless Sensor Networks

Swathi.Y, Sanjay Chitnis

Abstract : Increased demand for wireless communication system has gained huge attraction from various research communities, industries and academic field due to their significant advantages of facilitating efficient communication for real-time applications. With respect to wireless communication, monitoring (i.e. environmental) using WSN is considered as a significant task that has numerous challenging issues such as network deployment, data aggregation and data transmission in hazardous environmental conditions. In this work, we have focused on the data aggregation process in WSN and developed a novel approach to improve the network lifetime using a proposed solution based on the fuzzy logic scheme, which is called as Energy Aware Fuzzy Logic Secure Data Aggregation (EA-FSDA). Furthermore, we present a privacy-aware mechanism for secure data aggregation to improve the system reliability. The privacy-preserving scheme is developed using homomorphic data encryption scheme. Hence, the proposed approach provides a complete solution for efficient and protected data aggregation for WSN that aids in improving the performance of the network. Finally, we present a comparative study which proves that the proposed EA-FSDA technique attains improved network performance in contrast with existing techniques.

Keywords: Network security, wireless sensor network, fuzzy logic, clustering, encryption

I. INTRODUCTION

Recently, the demand for wireless communication has increased drastically in several real-time applications. The wireless communication systems include different types of communication systems such as cellular communication which are mainly used for voice communication and sensor network based communication which are used for monitoring the specific region [1]. Recently, wireless sensor network based communication systems are being adopted widely for real-time applications in industrial (i.e. monitoring the manufacturing are, development of various sensing devices etc.), military (i.e. land area monitoring, battlefield surveillance etc.) and academic purpose [2]. The extensive use of these type of networks urges for continuous development and improving the performance of communication systems.

Revised Manuscript Received on August 30, 2019.
* Correspondence Author
Swathi.Y*, Dept.of ISE1, CMR Institute of Technology, Bengaluru, India.
Sanjay Chitnis, Dept.of CSE2, Dayananda Sagar University, India.

Hence, communication in sensor network has fascinated research community because of its several significant characteristics such as reconfigurable architecture and efficient monitoring in hazardous environments [3]. We mainly focus on wireless sensor network communication and present a fresh method for improving the overall communication performance.

A. Wireless sensor networks

A WSN is produced using numerous sensor devices which are called as sensor nodes. These nodes come with restricted power supply batteries, data sensing, processing, and transmitting capabilities which are used for monitoring the desired field [4]. The technological advancements in the electronics field have enabled the growth in the low power integrated electronic devices, Micro-Electro-Mechanical systems (MEMS), and various computational technologies. This growth has improved the working efficiency of sensor networks, hence, these networks are widely adopted in various types of applications [5].

In general, WSNs are deployed randomly in a geographically remote area for monitoring the ecological as well as physical conditions of the region such as temperature, humidity, lighting conditions, pressure, chemical aspects, and fire detection. According to the working of the sensor network, information is collected by the sensor nodes and forward to base station for further processing. A sensor node contains actuators, central units (which is composed by using CPU and memory etc.) and Communication modules (RF communication systems). The complete process of sensor network based communication is mainly divided into two stages: (a) data collection [6] and (b) data packet routing towards base station (BS) [7].

Data collection is processed to gather the various type of data from different sensors and store in the sensor memory before transmitting to the corresponding sensor node. Similarly, the collected data packets need to be sent to the BS (Base Station) for further processing to take the appropriate action for the environment. According to the data transmission process, efficient path selection, shortest path identification, node clustering, and cluster head selection etc. tasks are performed. In [8] authors have discussed several advantages of efficient data aggregation such as it can minimize the no. of communication by reducing the redundancy which can be helpful for reducing the network energy usage, moreover the collision probability also can be reduced using efficient data aggregation scheme. Due to these advantages, several schemes have been introduced for data aggregation such as Ant Colony-based data aggregation [9], Croce et al.

[10] have introduced a novel approach for data aggregation for power consumption reducing using Fuzzy Logic based decision-making process. Wang et al. [33] focused on cluster head selection using fuzzy logic based on various parameters such as node's remaining energy, the centrality of the clusters as well as the distance between the node and the mobile sink.Xu et al. [11] extended this approach of data aggregation for multi-hop sensor network architecture and improved the latency performance. Similarly, Tang et al. [12] developed data aggregation protocol for mobile sensor nodes for extending the lifetime of the network. Conversely, data packet routing and transmission is additionally a vital parameter which can affect the communication performance. Several energy-aware routing protocols have been discussed in this field of WSN which are mainly divided into hierarchical and geographical routing protocols in WSN [13]. Cheng et al. [14] developed a novel hierarchal routing approach for network performance enhancement. According to this routing protocol base station information, node distribution density, distance from the sensor node and residual energy-related parameters are considered which helps to select the suitable routing protocol and its parameter for improving the network lifetime and energy consumption performance. Hua et. al. [15] developed combined routing and data aggregation approach for network lifetime improvement by conjointly optimizing the data aggregation as well as message routing. A scheme was presented to integrate the routing and data aggregation along with the optimization strategy. In this field of hierarchical routing protocol, LEACH [16], PEGASIS [17], TEEN [18], and APTEEN [19] etc. have been introduced which are used to improve the overall lifetime of the network. However, the first phase of the WSN is to aggregate the data efficiently which can be useful for improving the further process. In the proposed study, we mainly concentrate on improving the lifetime of the network and reducing the data redundancy. In data aggregation, privacy preserving and data security are also considered an important parameter. Othman et al. [20] presented data confidentiality and integrity based novel protocol for WSN using homomorphic encryption scheme. Similarly, Boubiche et al. [21] efficient watermarking based security scheme using fragile watermarking scheme for sensor networks. Hence, security and privacy preserving are also considered as an important parameter for improving the network lifetime performance.

B. Work contribution

As discussed in the previous section, the efficient data aggregation and security provisioning are an important parameter for improving the network lifetime and secure communication. Hence, in the proposed work, our key objective is to develop a fresh approach for energy efficient data aggregation where efficient cluster head selection, data encryption model and location privacy is developed.

The first phase focuses on the identification of the optimal number of cluster formation, in the second phase, cluster formation and cluster head selection scheme is developed. Later, we present data encryption and decryption strategy and finally, location privacy scheme is developed.

C. Article organization

The complete article is arranged in following manner: section II of the paper discusses about recent studies on secure data aggregation models, section III presents a proposed solution for the energy-aware and secure data aggregation models. Experimental study and performance comparison using the proposed model is presented n section IV and lastly in section V, final remarks and future direction for data aggregation in WSN are presented.

II. LITERATURE REVIEW

In this section, we study the prevalent and standard mechanisms in the field of WSN which are mainly focused on the data aggregation, security provisioning and privacy-preserving for sensor nodes. the efficient data aggregation is a promising aspect which can significantly improve the network performance. In this process of data collection in WSNs, both, maintaining data fidelity and energy efficient communication are considered as a challenging task. Several approaches have been introduced to make data collection more efficient and reliable. Xiang et al. [22] developed a novel approach for aggregation of data in WSNs using compressed sensing scheme which is targeted for improving the data reliability and energy efficiency during communication in random WSN topology. In order to achieve this objective, authors developed a wavelet-based method which provides a sparse basis to characterize the spatial and temporal correlation in a given random WSN topology. This scheme enables compressed sensing based data aggregation and high-reliability at the receiver end. In this work, the minimum-energy consumption process is formulated as NP-completeness problem which later solved using mixed integer programming along with the greedy approach. During data aggregation process in a system where higher density of nodes are present, chances of redundancy in the collected data is higher which in turn may degrade the information and can cause poor monitoring. Moreover, this data redundancy can cause energy consumption hence data fusion schemes can be incorporated for energy saving mechanism. According to this process, data aggregation can take place at the intermediate nodes which can help to reduce the size of collected redundant data resulting in the decreased cost and energy consumption. In order to achieve this, a novel data routing is presented for data fusion during intermediate node processing which helps to reduce the messages, increases aggregation rate and maintains trustworthy data aggregation and communication [23]. In this field of data aggregation in WSNs, clustering based approaches are widely adopted which shows a significant improvement in the data aggregation process due to significant nature of sensor nodes to provide reliability in clustered communication. Based on these assumptions of clustering Yuan et al. [24] presented spatial-correlation based method for data aggregation. In general, spatial-correlation based methods have been used for data aggregation in sensor networks [22] which can be useful for information fusion in the sensor networks.

However, authors have concluded that spatial-correlation based methods fail to perform efficiently in complex environments.

Moreover, these methods provide an inaccurate measurement for the real-time scenario.

These issues occur due to the inappropriate correlation degree between the current node and its neighboring node. In order to solve this problem,

an interesting approach was developed in [24] which provides the highest correlation degree between the sensor node's data using data density correlation degree based clustering technique. This technique helps to reduce the distortion and redundancy in the data. For energy-aware communication in WSN, LEACH protocol is considered a promising solution but due to the lack of high degree of data aggregation, this scheme shows poor performance for WSN data aggregation. Hence, Arumugam et al. [25] developed a novel approach, called energy-efficient LEACH (EE-LEACH). This process helps to collect the effective data using optimal clustering methods in this process, the sensor node clustering scheme is applied and every cluster chooses its cluster head. The cluster head selection process improves the energy dissipation performance and optimizes the resource utilization for data aggregation. Furthermore, a routing scheme is also presented which is established on the node residual energy i.e. the node which is having the highest energy is considered as the next Hop for packet transmission. During this course of aggregating data and transmitting it in hostile environments, usage of energy and life of network are two most important paradigms which can affect the overall communication performance. These issues can be addressed using clustering and appointment of cluster head (CH) for proficient data transmission [25]. The conventional approaches such as Low energy adaptive clustering hierarchy (LEACH) and EE-LEACH [25] uses probabilistic threshold based method for CH selection which can collect the information and transmit to the base station but due to excessive energy consumption, these methods fail to present an efficient solution for data aggregation and transmission in WSN. Hence, a super-CH (SCH) is selected based on the Fuzzy Logic based model for data forwarding. [26] The selection can be done based on the fuzzy rules which include remaining power, BS mobility, and cluster centrality. The study presented in [26] is based on the fuzzy logic where dynamic base stations are considered. In another study presented in [31], a Fuzzy Logic based approach is implemented where Fuzzy Logic controller considers energy, distance and the angle as an input and cluster head selection is done through Mamdani fuzzy inference system. Further, the A-Star search method is also incorporated to find the optimal route from the source node to the destination node. Khan et al. [34] developed Fuzzy-TPOSIS based approach for cluster head selection in WSN. In order to formulate the rules, the following criteria are considered which are as follows: node energy consumption rate; no. of neighbor nodes; left over energy; distance from the sink; and median distance between neighboring nodes.In order to further improve the performance of this approach, Yuan et al. [27] presented a fresh method for dynamic clustering using a genetic algorithm for presenting the self-organizing dynamic clustering. This approach is able to generate efficient modeling for cluster formation using predicted energy consumption, distance from the base station and node density. Tang et al. [28] developed trust-based routing for efficient and safe data collection in sensor networks. According to this approach, a signature based method is developed for data collection and maintaining the integrity in the network. Moreover, this process uses a sample packet transmission which contains node ID, data sending time-related information, to verify that whether the packet is successfully reaching to the destination node. By doing this, multiple paths are identified along with different trust values and the path which is having highest trust value is considered as the main path for transmitting the data packets. Securing the data and maintaining the privacy in nodes is assumed as a vital aspect for improving the performance of WSN. Boubiche et al. [21] presented a watermarking approach called SDAW (secure data aggregation watermarking-based scheme inhomogeneous WSNs). According to this scheme, the collected data is watermarked using a lightweight fragile watermarking scheme. This study shows that data encryption and decryption requires more time. Moreover, the communication links between communicating nodes are also watermarked to ensure the security. Similarly, Gopikrishnan et al. [29] developed a secure data aggregation approach for providing secure data aggregation in WSN. This scheme uses high secure data aggregation along with energy-efficient communication. in this work, the authors developed a secure authentication approach using a public key based method where the asymmetric cryptography scheme is presented for data integrity and confidentiality. In this section, we have studied the various approaches for efficient data aggregation where data integrity, quality, and security has an important role to make communication performance better. The sensor nodes are generally equipped with limited energy resources, hence efficient data aggregation is a recommended technique to improve the network lifetime. Moreover, maintaining privacy and security in the networks is also considered as an important parameter. Several schemes have been introduced to address these issues and still, these areas are considered as an important research field which can improve the overall performance of the sensor networks.

III. PROPOSED MODEL

We present the proposed solution for the energy-aware and secure data aggregation for WSN. The complete section contains the following sections as (a) Network Modeling and problem formulation, (b) Clustering & cluster head selection and (c) privacy & security preserving for the data aggregation. Figure 1 shows a complete process of the proposed approach. This article is arranged in 5 main sections as depicted in figure 1.

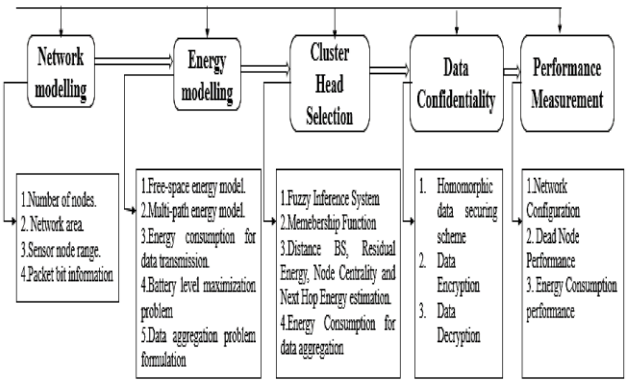

Fig.1 Flow chart of the proposed approach

A. Network modeling and problem formulation

We discuss about network modeling and later present the energy consumption minimization and data aggregation problem. Let us consider that a multi-hop WSN is expressed in the form of an undirected graph as $\mathbb{G} = (\mathbb{N}, \mathbb{A})$ where $\mathbb{N}$ represents the set of multiple numbers of nodes as $\mathbb{N} = \{1,2, \dots \mathcal{N}\}$, where $\mathcal{N}$ denotes the total number of nodes present in the network and $\mathbb{A}$, denotes the feasible number of arcs or links among nodes in the deployed network. Here we assume that each node has a fixed transmission range r. According to this model, initially, nodes are distributed randomly in region $\mathcal{R}$ where $\mathcal{R} \subset \mathbb{R}^2$. For generality, we consider that these sensor nodes are deployed in the square field where finite number of cluster head can be present in the region $\mathcal{R}$. In the given region $\mathcal{R}$, the distance between sensor node $i \in \mathcal{N}$ and $j \in \mathcal{N}$ is expressed as d_{ij}, distance from the current sensor node $i \in \mathcal{N}$ to base station BS is denoted by δ_i, the battery level of this node is denoted by b_i which helps to transmit a data packet of l bits size. During the transmission of l bit packet, transmitter and receiver circuits consume specific energy denoted by E J/bit, data aggregation coefficient is denoted by D_{DA} J/bit. In this work, we present a novel approach where each node is evaluated whether it has the positive level of batter for communication and total nodes which are having a positive energy level, are denoted as n and nodes are denoted as 0 if it doesn't have enough energy level otherwise 1. In order to determine the cluster head selection, we use a constant ϕ which is obtained using Fuzzy decision rules as discussed in next sub-section.. At this stage, data transmission is defined based on two models by considering the distance between sensor nodes similar to the LEACH protocol where a distance threshold is assumed as d_{th}, if the distance between sensor nodes is less than d_0 then free space modeling is applied otherwise multi-path modeling is applied for energy consumption. N free space model, energy consumption is proportional to the squared distance whereas, in multi-path model, the energy consumption is proportional to the biquadrate distance. Data transmission from node i to j requires a specific amount of energy which is expressed as:

$$D_{ij} = \begin{cases} -E + d_{ij}^2 \epsilon_{fs}, & (if\ d_{ij} < d_{th}) \\ -E + d_{ij}^4 \epsilon_{mp}, & (if\ d_{ij} \geq d_{th}) \end{cases} \quad (1)$$

Similarly, energy consumption for data transmission from source node i to base station BS is computed as:

$$F_i = \begin{cases} -E + f_i^2 \epsilon_{fs}, & (if\ f_i < d_{th}) \\ -E + f_i^4 \epsilon_{mp}, & (if\ f_i \geq d_{th}) \end{cases} \quad (2)$$

Where ϵ_{fs} (pJ/bit/m^2) denotes the free-space energy consumption constant and ϵ_{mp} (pJ/bit/m^4) denotes the energy consumption for multipath modeling. Similarly, in order to receive the data from a sensor node the amplifier energy consumption is denoted by lE.

Here, our first objective is to improve the energy consumption performance using a clustering problem formulation in a linear integer programming problem. The energy consumption minimization problem can be expressed as:

$$\text{maximize} \sum_{i \in \mathcal{N}} \left\{ b_i - \left(l \sum_{j \in N} D_{ij} y_{ij} + l F_i x_i \right) - lE \sum_{j \in \mathcal{N}} y_{ji} - lE_{DA} \sum_{j \in \mathcal{N}} y_{ji} \right\}$$

$$\text{Subject to } x_i + \textstyle\sum_{j \in N} y_{ji} + S_i = 1,\ i \in \mathcal{N} \quad (2$$

$$\left(b_i - \frac{\phi}{n} \sum_{k \in \mathcal{N}} b_k \right) x_i \geq 0, \qquad i \in \mathcal{N}$$

$$y_{ij} \leq x_j, \qquad i,j \in \mathcal{N}$$

$$x_i \in \{0,1\}, \qquad i \in \mathcal{N}$$

$$y_{ij} \in \{0,1\}, \qquad i,j \in \mathcal{N}$$

Where x_i is used for representing whether the node is selected as a cluster head if node i is selected in cluster head then it is represented using $x_i = 1$ otherwise $x_i = 0$. Similarly, y_i is denoted that whether node i belongs to the cluster where CH is selected. The main objective of this process is to maximize the sum level of all nodes battery levels after performing each round of communication.

Along with energy consumption minimization, we focus on data aggregation function which is defined as $y(t) \triangleq f(\partial_1(t), \partial_2(t), .. \partial_{\mathcal{N}}(t))$ where $\partial_i(t)$ denotes the reading of sensor i at time t. The given function f includes average, sum, minimum, maximum and count. Huge networks generate data traffic at high rates which becomes a tedious task to process at the server, moreover, it consumes more bandwidth and power hence data aggregation schemes are required for energy and power resource optimization. the additive data aggregation function can be expressed as:

$$f(t) = \sum_{i=1}^{\mathcal{N}} \partial_i(t) \quad (4)$$

Hence, we have two main objectives as power consumption minimization and improving the data aggregation efficiency to improve the overall network performance.

B. Cluster formation and cluster head selection

In order to improve the performance of the network, we present a Fuzzy Logic based modeling for cluster formation and cluster head selection. Fuzzy logic is an intelligent scheme which is also considered as non-linear data mapping approach. According to this process, the input data vector is provided output is obtained in terms of linguistic scalar

$$\begin{aligned} &\psi_{Member_1^i}(var_1') * \psi_{Member_2^i}(var_2') \\ &* \psi_{Member_3^i}(var_3') * \dots \\ &* \psi_{Member_q^i}(var_q') \\ &= T_l^q \psi_{Member_1^i}(var_1') \end{aligned} \quad (5)$$

form. In order to obtain the desired output, Fuzzy rules need to be defined which are used for making the decision and the complexity in decision making depends on the input parameters and fuzzy sets. Once the fuzzy rules are defined, the fuzzy inference system is applied which helps to generate a single fuzzy set to make the decision for the uncertain and imprecise information.

The system architecture of the Fuzzy Logic Model is depicted in below-given figure2.

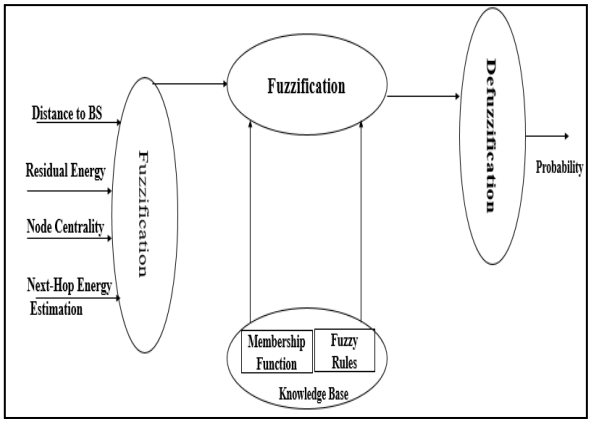

Fig.2. Fuzzy Logic Architecture

The main components of fuzzy logic are called as fuzzifier which helps to convert the crisp value into membership degree by using the corresponding membership function. The membership functions are used for determining the relationship between crisp value and its corresponding specific linguistic value.

Inference system: this component is used for making the decisions based on the which are formulated using different linguistic variables and these linguistic statements are stored in a rule base. These fuzzy rules are formulated in the form as "Premise: x is A., Implication: IF x is A THEN y is B and Consequent: y is B", these rules are composed of input variables which are connected using logical function (AND, OR, NOT) and the Consequent is an output variable. Let us consider that a q input 1-output Fuzzy Logic system is given as

R: Var_1 is $Member_1^i$ and Var_2 is $Member_3^i$ and … Var_q is $Member_q^i$ then the output is O'. Using Fuzzy logic, for input $Var' = \{var_1', var_2', var_3', \dots var_q'\}$, the degree of fuzzy rules provide the output as:

ψ denotes the membership function, $*$ denotes triangular membership function and T denotes the trapezoidal membership norms which are the binary operations applied to the fuzzy sets for membership function.

Defuzzification: this is a process which is used for obtaining the crisp values from fuzzy sets. Several methods are present for defuzzification of the fuzzy sets such as the center of gravity, maximum method, and center of a singleton. These methods are used for identifying the output based on the maximum activity of fuzzy sets and it provides the probability of output by computing the centroid which can be given as:

$$Centroid = \frac{\sum \psi_n(a) * a}{\sum \psi_n(a)} \quad (6)$$

At this stage, our aim is to formulate and cluster and selection of cluster head from the cluster members based on the following parameters which are: (a) distance to base station, (b) residual energy, (c) node centrality (c) expected energy consumption.

(a) **Distance to BS:** this is the measurement of the distance between node i and base station. According to the relationship of this, the energy consumption increases if the distance between node and base station. Minimization of distance between CH and BS can help to improve the energy saving performance.

(b) **Residual energy:** It is a measurement of remaining energy in the node after finishing the packet transmission to the cluster head. This is considered an important parameter for CH selection because low residual energy may cause node failure due to insufficient energy.

(c) **Node centrality:** node centrality is a measurement factor which shows how a node is placed at the center of the neighboring node among its neighboring nodes. if the node centrality is more for the particular node then it has more chances for CH selection. The node centrality can be computed as:

$$N_C = \frac{\sqrt{\sum_{i=1}^{N_D} \frac{dist_i^2}{N_D}}}{Network\_Dimenstion} \quad (7)$$

Where N_D denotes the degree of the node which represents a total number of neighboring node in the communication radius.

(d) Next-hop energy consumption estimation: this the measurement of required expected energy consumption to transmit the data packets to the next hop. If energy usage requirement is more, then the particular node can be discarded from the consideration of next hop. Let us consider that total μ number of queries are being received in a time period t. The estimated energy consumption for transmission can be given as:

$$E_t = (e_d r^n + e_t) \quad (8)$$

Where e_d denotes energy dissipation for transmission, r denotes the transmission range, η denotes the path loss coefficient. Based on these parameters, the probability of cluster head selection is computed where several fuzzy linguistic variables are considered according to the aforementioned parameters. In this work, fuzzy linguistic variables are mentioned in table 1.

Parameter	Linguistic Variable
Residual energy	{Low, Medium, High}
Node Centrality	{Close, Reachable, Distant}
Distance to BS	{Nearby, Average, Far}
Next Hop Energy consumption	{Low, Medium, High}

The corresponding representation in terms of the trapezoidal and triangular membership function. According to the proposed scheme, {Low, High, Close, Distant, Nearby, and Far} follow trapezoidal membership function, whereas {Medium, Reachable, and Average} follow triangular membership function. Based on these parameters, figure 3 shows Fuzzy logic representation based on the residual energy parameters.

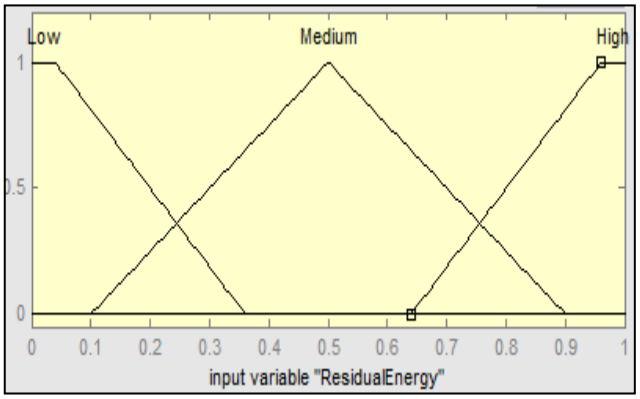

Fig.3. Residual membership function

Similarly, node centrality membership functions depicted in figure 4.

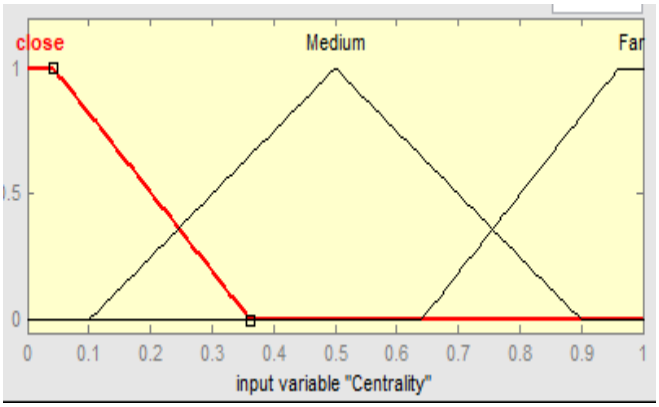

Fig.4. Node centrality membership function

Similarly, distance to BS and expected node energy consumption membership function are depicted in figure 5 and 6 respectively

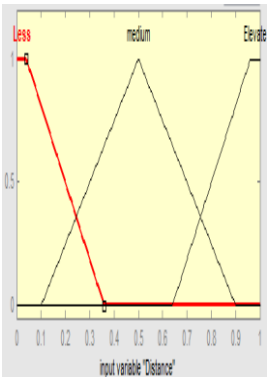

Fig.5. Distance to BS membership function

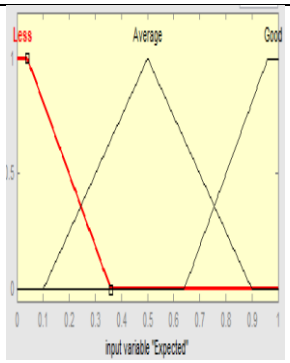

Fig.6. Expected energy consumption membership function

Based on these membership functions, the nine probabilities are obtained which are known as very low, low, moderately low, medium low, medium high, rather high, high, and very high. Here, very low and very high linguistic variables use trapezoidal membership function and the left out variables use triangular membership function. A simple if-then rule formulation is presented in table 2.

Table.2. If-then rules for Fuzzy Membership function

S. N	Residual Energy	Node Centrality	Distance to BS	Expected Energy	Probability
1	Little	Close	Near	Little	Rather Less
2	Little	Reachable	Near	Little	Little
3	Little	Far	Near	Little	Very Less
4	Little	Close	Normal	Medium	Rather Less
5	Little	Reachable	Normal	Medium	Little
6	Little	Far	Normal	Medium	Very Less
7	Little	Close	Distant	High	Rather Less
8	Little	Reachable	Distant	High	Little
9	Little	Far	Distant	High	Very Less
10	Moderate	Close	Near	Little	Rather Moderate
11	Moderate	Reachable	Near	Little	Moderate
12	Moderate	Far	Near	Little	Very Moderate
13	Moderate	Close	Normal	Moderate	Rather Moderate
14	Moderate	Reachable	Normal	Moderate	Moderate
15	Moderate	Far	Normal	Moderate	Very Moderate
16	Moderate	Close	Distant	Good	Rather Moderate
17	Moderate	Reachable	Distant	Good	Moderate
18	Moderate	Far	Distant	Good	Very Moderate
19	Good	Close	Near	Good	Very Good
20	Good	Reachable	Near	Good	Good
21	Good	Far	Near	Good	Rather Good
22	Good	Close	Normal	Moderate	Very Good
23	Good	Reachable	Normal	Moderate	Good
24	Good	Far	Normal	Moderate	Rather Good
25	Good	Close	Distant	Good	Very Good
26	Good	Reachable	Distant	Good	Good
27	Good	Far	Distant	Good	Rather Good

Once these membership functions are generated, the cluster formation scheme takes place where all nodes generate a arbitrary value in the range of 0 and 1. A predetermined threshold is considered for initial round and compared with the generated random value, if the generated value is higher than the threshold then all nodes under the communication range of this node form a cluster and this node are considered as a temporary CH node which is later determined using Fuzzy rules. After obtaining the appropriate CH, the members transmit data to the corresponding Cluster Head in the given time slots using TDMA model where data is aggregated sequentially. During this phase, the total energy consumption is given as:

$$Total_{Energy} = E_{intercluster} + E_{intracluster}$$
$$E_{intracluster} = E_{DA} + E_{Rx} + E_{member} \quad (9)$$

where E_{DA} denotes the data aggregation, E_{Rx} denotes the energy consumption for receiving the data packets and E_{member} denotes the energy consumption by cluster member to transmit the data which can be expressed as:

$$E_{member} = \sum_{i=1}^{k} \sum_{j=1}^{m} E_{Tx}(j, CH_i) \quad (10)$$

k denotes the optimal number of clusters obtained in the deployed network, m represents cluster members and $E_{Tx}(j, CH_i)$ represents the cost of transmission energy for transmitting the data from node j to cluster head. Similarly, total data aggregation energy consumption in the network is given as:

$$E_{DA} = m.n.E_{Singlebit} \quad (11)$$

Where n denotes the number of bits and $E_{Singlebit}$ denotes the energy consumption for aggregation of the single data bit. Inter-cluster energy consumption depends on the transmission of data from cluster head to the base station which can be expressed as:

$$E_{intercluster} = \sum_{i=1}^{k} \left(E_{Tx}(CH_i, BS)\right) \quad (12)$$

Based on these parameters, the total energy consumption for the cluster head by considering multipath loss scenario can be expressed as:

$$E_{CH} = n.E_{DA}\left(\frac{N}{k}\right) + n.E_{elec}\left(\frac{N}{k} - 1\right) + n.\varepsilon_{mp}.d^4 + n.E_{elec} \quad (13)$$

Similarly, energy consumption for cluster member in a free-space path loss (because the distance between member and head is minimum) can be expressed as:

$$E_{member} = n.\varepsilon_{fs}.d^2 + n.E_{elec} \quad (14)$$

After applying the energy-aware data aggregation scheme we present data confidentiality management scheme as discussed next sub-section.

C. Data Confidentiality

Here we present a cryptographic scheme for maintaining the data confidentiality using homomorphic encryption technique where algorithmic computation can be accomplished on the ciphertexts. Let us consider that $Enc()$ is a data encryption model where $\mathcal{M}$ denotes the message space and $\mathcal{C}$ denotes the ciphertext space which is under $\oplus$ and $\otimes$ operations respectively. This scheme is a homomorphic scheme as $(\oplus,\otimes)$ for any instance of $Enc()$ given as $c1 = Enck1(m_1)$ and $c2 = Enck1(m_2)$ where k represents the key such that $c1 \otimes c2 = Enc_k(m_1 \oplus m_2)$.

According to the homomorphic encryption scheme, the encryption process can be presented as:

$$Enc(m_1.r_1, p) = m_1 + r_1(mod\ p) \quad (15)$$

Where m_1 denotes the message to be encrypted, r_1 is the key used and p is the value used for computing the modulus. Moreover, this scheme adopts the additive homomorphic encryption process which can be given as:

$$\begin{aligned} &Enc(m_1.r_1, p) + Enc(m_2.r_2, p) \\ &= \left(m_1 + r_1(mod\ p) + m_2 \right. \\ &\left. + r_2(mod\ p)\right) \\ &= m_1 + m_2 + r_1 \\ &+ r_2(mod\ p) \\ &= Enc(m_1 + m_2, r_1 + r_2, p) \end{aligned} \quad (16)$$

In this work, we have considered a sample message as "$Transmit\ this\ over\ WSN$" which is aggregated and need to be communicated from source node to destination. With the help of encryption function, the encrypted message is obtained as "Ê-Ýqì$CI{\sim}FX$§)@" and the decrypted message is "$Transmit\ this\ over\ WSN$". According to this approach, the r_i keys are generated by the sensor along with the secret key (k_i) and unique node id for encrypting the message. The key sharing is performed among neighboring sensors which makes it more reliable for securing the information for a specific cluster. The data confidentiality scheme is adopted from [30]. Using this approach, we present a complete model for data aggregation which provides guaranteed confidentiality and integrity for WSN communication.

IV. RESULTS AND DISCUSSION

In this section, we demonstrate the experimental study using the proposed approach and evaluated the performance. The outcome of the proposed EA-FSDA (Energy Aware Fuzzy Logic Secure Data Aggregation) is compared with the existing protocols such as LEACH, EEUC, CHEF, MOFCA original and MOFC-optimized. In order to show the robust performance of proposed scheme we consider two test scenarios by considering the location of sink node such as (a) sink node is located far from the deployed area (b), sink node is located in the deployed region. These two scenarios are formulated to show the performance of the proposed approach for a varied distance of the base station from the network. The experimental work is implemented using MATLAB 2017a simulation tool running on a 2.70 GHz quad-core processor with 8 GB DDR4 RAM on Windows 8 OS with 1024 GB Drive. The complete network parameters are depicted in table 3.

Table.3. Simulation parameters

Parameters	Considered Values
No. of Nodes	100-400
Initial Energy	1J
Message Size	4000 bits
Control Message	200 bits
Sensing range	25 m
E_{elec}	50 nJbit^{-1}
ε_{fs}	10 pJbit^{-1}
ε_{mp}	0.0013 pJbit^{-1}

The performance of the proposed approach is compared in terms of a number of round vs the percentage of dead nodes. first of all, we consider that the network area is 100mx100m where BS is located at $(50,50)$ and a total number of 100 and 200 nodes are deployed in this region. once communication is initialized, we compute the FND (First Node Die) and HND (Half Node Die) statistics to measure the network performance.

Table.4. First Node Die And Half Node Die Performance Comparison

Algorithm	100 nodes		200 nodes	
	FND	HND	FND	HND
LEACH	335	769	300	743
CHEF	682	829	726	842
ECPF	538	831	479	743
EAUCF	636	796	685	810
MOFCA	640	785	683	784
FLECH	695	856	731	872
EA-FSDA	713	910	760	890

The comparative study shows that the proposed approach takes a number of rounds for first and half node dies for both 100 and 200 node scenario. This shows that the proposed EA-FSDA approach can be used for improving the network lifetime. A comparative study in terms of network lifetime for 200 nodes for this case (where BS is located at the center) is presented in figure 7.

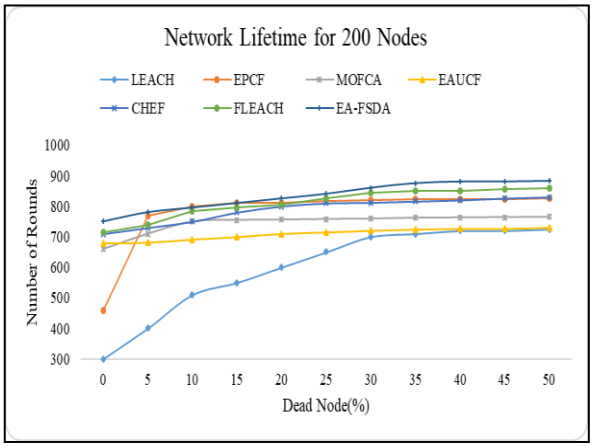

Fig. 7. Network lifetime performance for 200 nodes

Similarly, we evaluate the performance for 200 and 400 nodes where the base station is located outside the deployed region. for this experiment also, we have considered a similar setup and performance measurement parameters. With the help of the proposed approach, we obtain the performance of the proposed approach and compare with the existing techniques in terms of FND and HND as presented in table 5.

Table 5. FND and HND performance for 200 and 400 nodes

Algorithm	200 nodes		400 nodes	
	FND	HND	FND	HND
LEACH	10	161	17	180
CHEF	19	168	13	215
ECPF	8	115	7	197
EAUCF	13	160	14	218
MOFCA	13	215	13	281
FLECH	25	217	24	294
EA-FSDA	46	305	32	345

From table 5 it can be concluded that the proposed approach takes more round when compared to existing algorithms and hence the network lifetime can be prolonged for more number of nodes. The corresponding network lifetime performance comparison for 200 and 400 nodes is presented in figure 8 and 9 respectively.

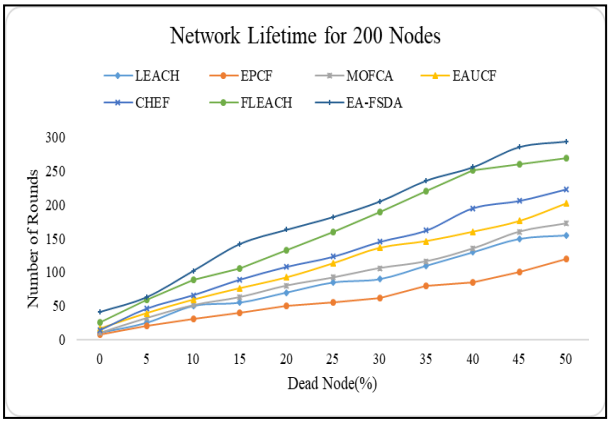

Fig.8. Network lifetime performance for 200 nodes (BS is located outside)

With the help of proposed approach, we obtain the average number of rounds as 85, 59, 92, 110, 125, 160 and 179 using LEACH, EPCF, MOFCA, EAUCF, CHEF, FLECH and EA-FSDA.

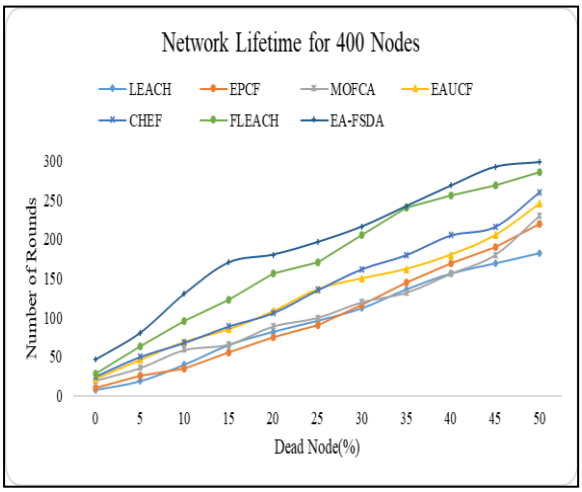

Fig.9. Network lifetime performance for 400 nodes.

Figure 8 shows a comparative performance in terms of a number of round vs dead node percentage. The average number of round performance is obtained as 97, 103, 108, 128, 136, 172 and 193 using LEACH, EPCF, MOFCA, EAUCF, CHEF, FLECH and EA-FSDA, respectively.

Based on these experimental scenarios, we present a comparative study in terms of energy consumption for scenario 1 where BS is located at the center coordinates and scenario 2 where BS is located outside. This performance is evaluated for 100 and 200 nodes as depicted in table 6.

Table.6. Energy consumption performance for 100 and 200 nodes (for the varied location of BS)

Algorithm	100 node		200 Nodes	
	BS at the center	BS outside	BS at the center	BS outside
LEACH	0.1137	0.26116	0.2180	0.4558
CHEF	0.1106	0.20882	0.2161	0.3225
ECPF	0.1106	0.2439	0.2169	0.35422

EAUCF	0.1151	0.1853	0.2285	0.30576
MOFCA	0.1177	0.18324	0.2331	0.3163
FLECH	0.10	0.1792	0.2151	0.3029
EA-FSDA	0.095	0.1017	0.1855	0.2658

The complete experimental performance analysis for a varied number of nodes for different locations of BS. The complete analysis shows that the proposed approach achieves promising performance when compared with the existing techniques. Similarly, to show the robustness of the proposed approach, we considered the 1000 number of nodes and evaluated the enactment in terms of network lifetime as depicted in Figure 10.

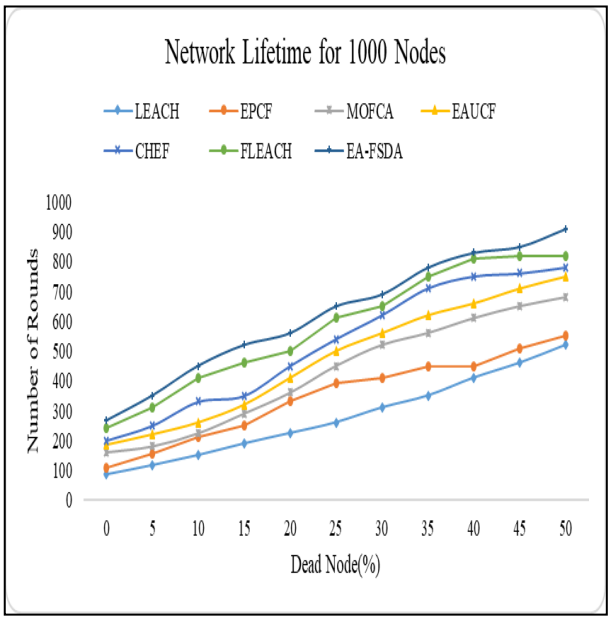

Fig.10.Network Lifetime performance for 1000 nodes

The above-given figure shows a comparative study in terms of a number of rounds (which represents the network lifetime). According to this experiment, the obtained number of rounds are 270, 347,426, 473,521,580, and 624 using LEACH, EPCF, MOFCA, EAUCF, CHEF, FLEACH and EA-FSDA.Moreover, the proposed solution of data security can be incorporated in real-time applications also such as temperature monitoring, securing ZigBee based communication (a study presented in [32] shows the failure of ZigBee based system because it received commands from multiple users), motor controller (study presented in [32] shows that inappropriate information about controller i.e. speed can be hacked which may lead towards the physical damage) in industrial applications [32].

V. CONCLUSION

In this work, we have focused on the significance of WSN and their related issues for data aggregation, privacy in aggregation and network lifetime. Several schemes have been introduced to overcome these issues in sensor networks but attaining the desired outcomes in terms of a network lifetime and energy usage along with data privacy is considered a challenging task. However, clustering schemes show a significant improvement in the data aggregation and lifetime enhancement purpose. In order to address these issue, we present a novel approach for data aggregation using a clustering scheme called as **E**nergy **A**ware **F**uzzy Logic **S**ecure **D**ata **A**ggregation (EA-FSDA) where Fuzzy Logic based scheme is used for efficient cluster head selection. Later, data privacy scheme is developed using homomorphic encryption based method. Finally, we present a comparative analysis which illustrates that the proposed mechanism attains higher performance while comparing other state-of-the-art methods.

REFERENCES

1. Zhang, J., Shan, L., Hu, H. and Yang, Y., 2012. Mobile cellular networks and wireless sensor networks: toward convergence. IEEE Communications Magazine, 50(3).
2. Othman, M.F. and Shazali, K., 2012. Wireless sensor network applications: A study in the environment monitoring system. Procedia Engineering, 41, pp.1204-1210.
3. de Amorim, M., Ziviani, A., Viniotis, Y. and Tassiulas, L., 2008. Practical aspects of mobility in wireless self-organizing networks [Guest Editorial]. IEEE Wireless Communications, 15(6), pp.6-7.
4. Bhattacharyya, D., Kim, T.H. and Pal, S., 2010. A comparative study of wireless sensor networks and their routing protocols. Sensors, 10(12), pp.10506-10523.
5. Guo, L., Fang, W., Wang, G. and Zheng, L., 2010, June. Intelligent traffic management system base on WSN and RFID. In Computer and Communication Technologies in Agriculture Engineering (CCTAE), 2010 International Conference On (Vol. 2, pp. 227-230). IEEE.
6. Di Francesco, M., Das, S.K. and Anastasi, G., 2011. Data collection in wireless sensor networks with mobile elements: A survey. ACM Transactions on Sensor Networks (TOSN), 8(1), p.7.
7. Senouci, M.R., Mellouk, A., Senouci, H. and Aissani, A., 2012. Performance evaluation of network lifetime spatial-temporal distribution for WSN routing protocols. Journal of Network and Computer Applications, 35(4), pp.1317-1328.
8. Akkaya, K., Demirbas, M. and Aygun, R.S., 2008. The impact of data aggregation on the performance of wireless sensor networks. Wireless Communications and Mobile Computing, 8(2), pp.171-193.
9. Liao, W.H., Kao, Y. and Fan, C.M., 2008. Data aggregation in wireless sensor networks using ant colony algorithm. Journal of Network and Computer Applications, 31(4), pp.387-401.
10. Croce, S., Marcelloni, F. and Vecchio, M., 2008. Reducing power consumption in wireless sensor networks using a novel approach to data aggregation. The Computer Journal, 51(2), pp.227-239.
11. Xu, X., Wang, S., Mao, X., Tang, S. and Li, X.Y., 2009, May. An improved approximation algorithm for data aggregation in multi-hop wireless sensor networks. In Proceedings of the 2nd ACM international workshop on Foundations of wireless ad hoc and sensor networking and computing (pp. 47-56). ACM.
12. Tang, S., Yuan, J., Li, X., Liu, Y., Chen, G., Gu, M., Zhao, J. and Dai, G., 2010, June. DAWN: energy efficient data aggregation in WSN with mobile sinks. In Quality of Service (IWQoS), 2010 18th International Workshop on (pp. 1-9). IEEE.
13. Al-Karaki, J.N. and Kamal, A.E., 2004. Routing techniques in wireless sensor networks: a survey. *IEEE wireless communications*, *11*(6), pp.6-28.
14. Cheng, H.B., Geng, Y.A.N.G. and Hu, S.J., 2008. NHRPA: a novel hierarchical routing protocol algorithm for wireless sensor networks. The Journal of China Universities of Posts and Telecommunications, 15(3), pp.75-81.
15. Hua, C. and Yum, T.S.P., 2008. Optimal routing and data aggregation for maximizing the lifetime of wireless sensor networks. IEEE/ACM Transactions on Networking (TON), 16(4), pp.892-903.
16. Heinzelman, W.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, HI, USA, January 4–7, 2000; pp. 1–10.
17. Lindsey, S.; Raghavendra, C.S. PEGASIS: Power-efficient gathering in sensor information systems. Proceedings of IEEE Aerospace Conference, Big Sky, MT, USA, 2002; pp. 1125–1130.
18. Manjeshwar, A.; Agrawal, D.P. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. 15th International Parallel and Distributed Symposium, San Francisco, CA, USA, April 2001; pp. 2009–2015.
19. Manjeshwar, A.; Agarwal, D.P. APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. Proceedings of the 16th International Parallel and Distributed Processing Symposium, Ft. Lauderdale, FL, USA, April 2002; pp. 195–202.

20. Othman, S.B., Bahattab, A.A., Trad, A. and Youssef, H., 2015. Confidentiality and integrity for data aggregation in WSN using homomorphic encryption. Wireless Personal Communications, 80(2), pp.867-889.
21. Boubiche, D.E., Boubiche, S., Toral-Cruz, H., Pathan, A.S.K., Bilami, A. and Athmani, S., 2016. SDAW: secure data aggregation watermarking-based scheme inhomogeneous WSNs. Telecommunication Systems, 62(2), pp.277-288.
22. Xiang, L., Luo, J. and Rosenberg, C., 2013. Compressed data aggregation: Energy-efficient and high-fidelity data collection. IEEE/ACM Transactions on Networking (TON), 21(6), pp.1722-1735.
23. Villas, L.A., Boukerche, A., Ramos, H.S., de Oliveira, H.A.F., de Araujo, R.B. and Loureiro, A.A.F., 2013. DRINA: A lightweight and reliable routing approach for in-network aggregation in wireless sensor networks. IEEE Transactions on Computers, 62(4), pp.676-689.
24. Yuan, F., Zhan, Y. and Wang, Y., 2014. Data density correlation degree clustering method for data aggregation in WSN. IEEE Sensors Journal, 14(4), pp.1089-1098.
25. Arumugam, G.S. and Ponnuchamy, T., 2015. EE-LEACH: development of energy-efficient LEACH Protocol for data gathering in WSN. EURASIP Journal on Wireless Communications and Networking, 2015(1), p.76.
26. Nayak, P. and Devulapalli, A., 2016. A fuzzy logic-based clustering algorithm for WSN to extend the network lifetime. IEEE sensors journal, 16(1), pp.137-144.
27. Yuan, X., Elhoseny, M., El-Minir, H.K. and Riad, A.M., 2017. A genetic algorithm-based, dynamic clustering method towards improved WSN longevity. Journal of Network and Systems Management, 25(1), pp.21-46.
28. Tang, J., Liu, A., Zhao, M. and Wang, T., 2018. An aggregate signature based trust routing for data gathering in sensor networks. Security and Communication Networks, 2018.
29. Gopikrishnan, S. and Priakanth, P., 2016. HSDA: hybrid communication for secure data aggregation in a wireless sensor network. Wireless Networks, 22(3), pp.1061-1078.
30. Di Pietro, R., Michiardi, P., and Molva, R., 2009. Confidentiality and integrity for data aggregation in WSN using peer monitoring. Security and Communication Networks, 2(2), pp.181-194.
31. Nanda, A., & Rath, A. K. (2018). Fuzzy A-Star Based Cost-Effective Routing (FACER) in WSNs. In Progress in Advanced Computing and Intelligent Engineering (pp. 557-563). Springer, Singapore.
32. https://www.electronicdesign.com/communications/secure-wireless-sensor-networks-against-attacks (Accessed on 27-Jan-2019)
33. Wang, J., Niu, J., Wang, K., & Liu, W. (2018, January). An energy efficient fuzzy cluster head selection algorithm for WSNs. In Advanced Image Technology (IWAIT), 2018 International Workshop on (pp. 1-4). IEEE.
34. Khan, B. M., Bilal, R., & Young, R. (2018). Fuzzy-TOPSIS based cluster head selection in mobile wireless sensor networks. Journal of Electrical Systems and Information Technology, 5(3), 928-943.

AUTHORS PROFILE:

Mrs.Swathi Y is presently pursuing Ph.D in computer Science and engineering in CMR Institute of Technology, Bangalore Karnataka. She obtained her Masters degree in computer Networks and engineering from CMRIT, Bangalore in the year 2012.She has published many research papers in national and International Journals and conferences. Her area of research includes game theory and Network Security.

Dr. Sanjay Chitnis is committed to mentoring students and faculty to pursue their life dreams and is passionate about transforming engineering education. He has over 22 years of vast leadership experience in multinational companies Motorola and LG. His expertise includes Software Product and Process Engineering, Program Management and Platforms and Apps for smart devices. He has a Ph.D. in Computer Science from IISc, Bangalore, M.Tech. in Electrical Engineering from IIT Kanpur and a B.E. in Instrumentation from University of Pune. Currently he is working in Dayananda Sagar University as Director, Innovation Cell.