An Efficient Routing Mechanism for Node Localization, Cluster Based Approach and Data Aggregation to Extend WSN Lifetime

Priyanka Rajanikanth[1]*   Kalli Satyanarayan Reddy[1]

[1]*Department of Information Science and Engineering, Cambridge Institute of Technology Bangalore, Affiliated to VTU, Belagavi (Karnataka), India*
* Corresponding author's Email: priyanka.89.r@gmail.com

Abstract: The Last decade the demand for wireless communication has increased massively. Wireless Sensor Networks (WSN) have turned out as a potential solution for various real-time applications namely environmental monitoring, health and military applications. These networks follow the random placement strategy hence, identifying the location of sensor nodes is a tedious task. Moreover, these networks suffer from various issues such as limited power resources. This work, focusses on the sensor node localization and network lifetime enhancement of Wireless Sensor Networks (WSN). In localization, security is an important issue, so a trust-based model to secure localization is investigated in this work. Further, energy aware clustering and Cluster Head (CH) selection has been undertaken to prolong the lifetime of the network. Finally, a data aggregation model which decreases the network energy depletion by discarding the redundant data is presented. The outcome of the model which is proposed has been assessed and expressed in the form of localization error, energy depletion and alive node. The comparative analysis shows that the localization error for varied anchor nodes is obtained as 1.51, 2.28, 2.52 and 1.25 using existing techniques DMA (Distance Mapping Algorithm), MDS-Map (Multidimensional Scaling), DV-Hop (Distance Vector-Hop) and Proposed Approach, respectively. Similarly, the average energy consumption is obtained as 1.02, 1.67, 1.90 and 0.82 using DMA (Distance Mapping Algorithm), MDS-Map (Multidimensional Scaling), DV-Hop (Distance Vector-Hop) and Proposed Hybrid Approach. The comparative analysis shows that proposed hybrid approach of localization, CH (Cluster Head) formation and aggregation attains better performance when compared with existing techniques in WSN(Wireless Sensor Networks).

Keywords: Clustering, Data aggregation, Energy consumption, Localization, Performance.

1. Introduction

The demand for wireless communication is increasing rigorously. The use of wireless communication has given rise to the evolution of new communication technologies. Currently, Wireless Sensor Networks (WSNs) have received a huge consideration in several real-time application scenarios. The WSN (Wireless Sensor Networks) is designed and deployed for specific applications and composed by various low-power and low-cost devices known as sensors. These sensors follow the ad-hoc scheme for node placement to perceive and transfer the application-specific data [1]. The sensor nodes are fortified with low-power devices (batteries),

and are generally installed in a hostile environment, where replacing the batteries is not a feasible task [2]. In the entire communication, sensor nodes consume excessive power which results in depleting the network energy. Hence, maintaining the network lifetime and deploying Sensor Network continuously is not a cost-effective solution. Hence, maintaining the network lifetime and minimizing energy consumption are the two serious issues in WSN communication [3]. Network lifetime maximization depends on several aspects of sensor nodes such as Data Aggregation, in this process, Sensor Nodes gathers the data and perform the aggregation process at the aggregator node. This process discards the redundant data and minimizes retransmission thus data aggregation plays important role in energy

consumption minimization [4, 5]. Clustering is another aspect where node clustering is an important phase where similar energy nodes are grouped together. In case, the sensor nodes with different energy constraints are grouped together then it affects the reliability and stability of the cluster. Moreover, the cluster head selection has to be performed in such a way that it can prolong the network lifetime [6, 7]. The next aspect being Routing which is the most important phase which transmits the data from the point where any specific event has occurred known as source node to the destination node by using multi-hop routing [8, 9]. Thus, energy-aware routing schemes are highly appreciated in this WSN.

In this article, improving the network lifetime has been focused by developing an efficient mechanism for data aggregation, cluster establishment, and cluster head selection. The newly developed approach considers trust-based localization, a dynamic cluster head selection and data aggregation. On the contrary, deducing the physical coordinates of sensor nodes is the major factor to identify the place of origin of events. This mechanism is known as sensor node localization. Several researches have been carried out for node localization as mentioned in [10, 11]. Achieving an efficient localization accuracy is a tedious task that needs to be addressed for improvement of event detection. Moreover, maintaining the secrecy about the senor node position and event is also an important aspect. Hence, the focus is on the secure node localization scheme in Wireless Sensor Networks.

These networks are deployed in dense and hostile environments where the actual position of these nodes is not realized. Global Positioning System (GPS) and Cellular Systems are considered as the most promising technique for exact positioning of nodes. Despite promising performance of the localization, installing a GPS (Global Positioning System) system for each node is not feasible due to higher implementation costs and energy consumption. The cellular positioning systems suffer from the deep shadowing effect [12]. Moreover, limited energy resources and implementation cost are the crucial parameters hence only some nodes which are known as beacon or anchor nodes, are allocated the Global Positioning System. Generally, the sensor node localization scheme is grouped into two different classes as range dependent and range independent localization. The range-dependent localization methods use distance and angle to find out the sensor nodes location. On the contrary, the range-free dependent technique use pattern matching scheme to estimate the location. As stated in [13, 14], several techniques have been identified to improve localization efficiency. Generally, the range-dependent localization schemes such as Time Of Arrival (TOA), Time Difference of Arrival (TDOA), Angle of arrival (AOA), and Received signal strength indicator (RSSI) are commonly adopted in localization. Hence, the focus is on the range-dependent localization scheme in this work. Moreover, the focus is on incorporating security during localization. Data aggregation is an important factor that has more effect on the quality of data (it discards the redundant data), and energy consumption. Several techniques are introduced for data aggregation schemes and these schemes are classified as cluster-based, centralized, tree-based, and in-network data aggregation. Similarly, cluster construction and cluster head selection techniques also affect the network lifetime. Hence, efficient cluster construction and selecting the cluster head is considered as an important step to improvise the network lifetime. Currently, hierarchal clustering schemes have gained attraction in Wireless Sensor Networks which are further classified as cluster-based and grid-based techniques [15]. However, these schemes fail to achieve the desired performance for huge network scenarios.

From this discussion, it can be concluded that secure data aggregation, cluster realization, and selecting the cluster heads helps us as the first step to proliferate the network lifetime. In this work, the objectives and development of a new methodology to deal with the network lifetime related problems has been focused. Limited work has been carried out, focusing on these issues. This article offers several mechanisms to boost the overall network performance. These key findings are, a secure localization scheme, where trust-based method is applied before localization is presented. The localization scheme is based on the RSSI (Received Signal Strength Indicator) technique Cluster realization and Cluster Head selection method which considers residual energy as the main parameter is presented. "Average" function-based aggregation of data to abate the energy depletion of the network is presented.

The work has been presented in three sections. In Section 2 under materials and methods, literature review is presented on secure localization, data aggregation, cluster realization and cluster head selection also it describes the proposed solution for the aforementioned objectives, in Section 3 the results and comparative analysis is presented and finally, in Section 4 the concluding remarks about the approach is presented.

2. Literature review

This section concisely describes the existing techniques which includes secure node localization, data aggregation, cluster realization and cluster head selection.

2.1 Secure node localization

The authors in [16] reported that existing techniques postulate the model parameters to decrease the complexity in cost functions in the process of position estimation. Although, these parameters depend on the functional conditions that prevail in wireless environment and adapt the changes over time due to obstacles in the environment and other variations in the environmental conditions. Performance of these systems cannot be guaranteed in real-time systems. To address these issues, challenges, authors introduced a combined model, where RSS (Received Signal Strength Indicator) and TOA (Time of Arrival) techniques which can estimate and adapt these challenges "on field". This scheme is split into two distinct steps, the first stage exploits the nodes in known position and estimates the unidentified parameters of Received Signal Strength Indicator and Time of Arrival. In next stage, hybrid TOA-RSS(Time of Arrival-Received Signal Strength Indicator) pooled with least square method to assess the position of unknown target. The authors in [17] modified the range-based localization algorithm by incorporating the use of distributed computing as cooperative and collaborative manner.

This approach focuses on the resource utilization and improvement in the position estimation. According to this scheme, the Media Access Control scheme (MAC) is studied as part of localization concept. The Media Access Control model helps to decrease the energy depletion by reducing the packet retransmission and collision. Complete network is divided into several clusters, where Media Access Control scheme restricts the coverage to limited neighbors. The authors in [18] have used triangulation method to estimate the node location at single hop distance from an anchor node. However, localization in terms of multi-hop is also introduced which has gained attention because of its increased distance of localization such as DV-Hop (Distance Vector Hop). DV-Hop(Distance Vector Hop) is widely adopted due to less complexity and implementation cost. But, Distance Vector Hop technique suffers from the issue of localization accuracy. To overcome this issue, authors introduced improved Distance Vector Hop approach which uses RSSI model values of links between one-hop

neighbors. Similar to this hybrid scheme, the authors in [19] presented hybrid approach using AOA/RSSI because these techniques do not require any synchronization among anchor and unknown nodes.

The conventional AOA schemes require large number of antennas which creates additional complexity in the network environment. To overcome these issues, authors presented 1AOA/n RSSI (Received Signal Strength Indicator) to increase the performance of Wireless Sensor Networks. Unlike previous techniques, this scheme requires one master anchor to obtain the location estimation using AOA (Angle of Arrival) whereas other anchor nodes are equipped with single-antenna transceivers. Initially, the AOA/RSSI (Angle of Arrival/ Received Signal Strength Indicator) models are realized for single virtual anchor further it is extended for 1AOA (1Angle of Arrival) and n RSSI(Received Signal Strength Indicator) models using least square and subspace methods. The authors in [20] focused on the localization accuracy problem of fixed multi-anchor methods and presented Three-Dimensional (3D) localization algorithm by combining RSSI (Received Signal Strength Indicator) and TOA(Time of Arrival) for single mobile anchor node scenario. In a 3D (3 Dimension) localization solution, a mobile anchor node is used, which travels in accordance with Gauss-Markov 3D mobility model [20]. The combination of RSSI (Received Signal Strength Indicator) and TOA (Time of Arrival) is implemented for near end and far end ranges, respectively to achieve the accurate gap between anchor and unidentified node. Finally, these values are used in maximum-likelihood estimation to obtain the positions of unknown nodes. However, this method doesn't deal with the mobility and multiple anchor node scenario, thus it is the main consideration for future research scope. The authors in [21] presented a hybrid model by combination of metaheuristic optimization like genetic algorithm, and particle swarm optimization. Further, these models are processed through the back propagation neural network to uplift the network performance. The authors in [22] presented dragonfly–firefly algorithm (DA–FA) for localization. The authors in [23] developed two-stage particle swarm optimization algorithm for estimation of node's position in the "concave regions". The complete process is comprised of two stages where first stage is used to measure the gap between nodes where it uses similar path search mechanism along with intersection ratio based on preliminary localization. In second stage, improved PSO (Particle swarm optimization) is applied on the outcome of initial localization. Although, this approach is suitable to

deal with positioning issues but overcoming the practical limitation of deploying the anchor node still remains a challenging task. Security is also considered as a prime concern in localization. Several schemes are presented to incorporate the security in the Wireless Sensor Networks localization. The authors in [24] considered a scenario of node which is malicious which tries to interrupt the procedure of localization Wireless Sensor Networks and presented the secure localization process which uses the Least Square method which is weighted (WLS) and the secure version of it (SWLS), LN-1E and LN-1 (ℓ1-norm based localization techniques). The WLS (weighted least square) and SWLS (secure weighted least square) techniques identify and eradicate the measurement of hostile node. In the case of planned attack, the LN-1E (Norm Based) technique models the problem as fitting the plane problem in which the computation of genuine and malicious node is separated by different plane. The computational complexity of this scheme depends on the number of anchor node thus increasing the anchor node leads to increased computational complexity.

2.2 Data aggregation

Data aggregation is also considered as vital parameter that can improve the energy depletion of the network. This section presents a discussion on data aggregation techniques. The authors in [25] focused on dealing with network lifetime related issue, WSNs have adopted duty-cycling technique where sensing capabilities of sensor nodes are switched on and off periodically. However, these techniques affect the data aggregation (DA) performance. To overcome this issue in DA, authors have introduced an approach called delay efficient data aggregation scheduling (DEDAS-D) technique in duty-cycled WSNs. The authors in [26] focused on securing the data aggregation mechanism and presented energy-aware secure data aggregation. However, this technique achieves the accuracy similar to maximum likelihood but achieving less error for highly dense networks is a challenging task for this approach. The authors in [27] introduced ring-based data aggregation scheme to boost up the energy efficacy of the network. This approach divides the network into multiple rings and data aggregation is performed through consequent rings from outside to insider direction. This approach focuses on both transmission reliability and energy consumption issues. For reliable transmission, the intermediate or aggregator nodes unicast the multiple copies however, more number of copies creates auxiliary energy cost. However, the energy utilization is optimized by adaptively selecting the packet copies using fuzzy logic. The authors in [28] emphasized on security along with energy aware data aggregation in Wireless Sensor Networks. The authors have evaluated a continuous data aggregation issue which is affected due to selective forwarding attacks. This model embraces the security by considering Shamir's sign and secret sharing scheme for sharing the data secretly considering the signatures for secure aggregation. The authors in [29] have used optimization strategy to extend the network lifetime with the help of firefly optimization. Moreover, authors have reported that proper Cluster Head selection is also considered as promising technique which can minimize the energy depletion. The authors in [30] have focused on improving the data aggregation process along with security concerns in the Wireless Sensor Networks (WSN). This scheme uses a novel secure data aggregation scheme where autoregressive model is used for time series analysis. This is a time series analysis technique which is used to avoid the learning of data pattern by adversaries.

2.3 Clustering and cluster head selection

The formation of the clusters and selecting the cluster head is the third objective which is a key factor in network QoS (Quality of Service) and lifetime. Generally, the node that has the higher residual energy is being considered as the best solution for CH (Cluster Head) selection. The authors in [31] have proclaimed that appropriate CH (Cluster Head) selection can be useful in load balancing in the network resulting in the enhancing the network lifetime. The authors presented a dynamic way for cluster head selection where CH (Cluster Head) is chosen depending on the variation of energy. This approach takes account of the energy in the initial state, remaining energy and CHs which has optimum values for the upcoming Cluster Head selection for specific clusters. Authors in [32] developed a cluster formation and selection of Cluster Head mechanism to improvise the network in terms of the energy efficiency by utilizing the optimization strategies. In this approach, authors introduced a new fitness function for hybrid model of Glow-worm swarm with Fruit fly Algorithm (FGF) [32]. The authors in [33] have presented an energy-aware routing using threshold and fuzzy logic-based approach. This model computes the probability value of cluster head selection by taking residual energy into consideration for each node. The data to forwarded to the sink by the cluster head with the implementation of fuzzy control mechanism with communication with multiple hops which makes use of length of the queue,

Table 1. Notation used in this work

Notation	Representation
Sensor nodes	S
Network region	$m \times m$
Unknown nodes	U_n
Coordinates of nodes	Ω
Path loss vector	Y
Probability function	P
Comprehensive trust	C
Weight factor	α and β
Comprehensive trust	C
Nodes attribute set	A
Time decay function	T
Direct trust between point x and y	D_{xy}
Recommended trust	R
Trust value of mediator node	M
Entropy	H
Energy consumption	E

the remaining energy and distance between node and base station. The authors in [34] have developed a "Enhanced Technique based on Reliability for Ordering the Similarity Preference by to Ideal Solution (RE-TOPSIS)" approach which uses the logic based on fuzzy method for decision making for cluster head selection. It selects the initial cluster and cluster head using LEACH (Low-energy adaptive clustering hierarchy) protocol further RE-TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution), removes the phase of selecting the cluster head in each round. Authors have considered several benchmarks such as the energy which is remining, the node distances, energy consumption rate, neighbouring node, and distance that is considered between cluster head and sink, distance between Cluster head and the node which is a member, and reliability index. The authors in [35] have developed a high-quality clustering algorithm to increase the intra and inter-cluster distance which reduces the clustering error. To select the cluster head which is optimal, the fuzzy logic evaluates several criteria like the remining energy, min and max distance which is in-between base station and the sensor node, min and max energy in each cluster.

In [36] Rong Tan et al. focused on improving the localization and presented Distance Mapping Algorithm (DMA) to identify the node position with the help of estimation and distance matrix with an optimized linear transforming function. Further, genetic algorithm is also incorporated to minimize the localization error. The o objective function in GA suffer from the poor convergence performance.

Salim El Khediri et al. [37] presented K-means algorithm to deal with energy consumption related

issues in WSN and improving the network lifetime. This is achieved by incorporating a sample space partition in k-means. Based on this, it selects the cluster head based on the network size. This scheme is known as optimal K-means however this scheme doesn't consider aggregation and security aspects.

3. Proposed model

In this segment a proposed solution for Secure Node Localization, Cluster realization, and Data Aggregation to boost up the network lifetime is presented.

3.1 Secure node localization

The first objective is to incorporate the security aspects in the node localization. Previous section described that range-based schemes are broadly adopted in this field, specifically, the RSSI (Received Signal Strength Indicator) algorithm reported significant improvement in the localization performance. Hence, Received Signal Strength Indicator approach is used for localization. In order to consider the security scenario, a trust evaluation-based model is presented to identify and avoid the malicious node.

3.1.1. Problem formulation

Consider that $S_l = \{S_1, \ldots, S_{n_a}\}$ is the set of sensor nodes that are heterogeneous in nature and which are installed arbitrarily in a two-dimensional geographical area which is divided into $m \times m$ cells and $\Omega = \{1, 2, 3, \ldots, m^2\}$ is the sample space for cell coordinates. The first n_a nodes are considered as the landmark as $S_l = \{S_1, \ldots, S_{n_a}\}$ and the unknown nodes are denoted as $S_u = \{S_{n_a+1}, \ldots, S_N\}$, y_{ij}^l represents the path loss samples which are collected by S_j from S_i at time stamp l. Let Θ_k be a vector which stores the path loss samples of different nodes which have been communicated between connected node pairs during first k time steps and $Y_j^{(k)}$ is the vector of all path loss samples which are collected by S_j from its neighbouring node. Consider a random variable $\tilde{x}_j^{(k)}$ which is defined over Ω to represent the location of S_j at k^{th} time step. In order to estimate the location of S_j at time step M for given path loss, the location estimation problem as is defined as:

$$\tilde{x}_j = \underset{x_j}{\mathrm{argmax}}\left[P\left(\tilde{x}_j^{(M)} = x_j|\Theta_M\right)\right] \qquad (1)$$

Where $P(.)$ represents the probability function and $\arg\max_{xj}[(f(x)]$ is used to denotes the largest value of x achieved by the function $f(x)$. Table 1 denote the notations used in this work.

3.1.2. Localization scheme

Consider a Sensor Network which comprises of beacon nodes and regular nodes. It is assumed that beacon nodes are capable of self-positioning in the network using positioning devices whereas the regular sensor nodes need to localize their position based on the information obtained from other nodes and beacon nodes. According to this approach, initially, the information obtained from beacon node is used for estimating the location information by using signal attenuation formula. Maximum likelihood estimation method is then applied, when enough data is collected by the beacon nodes and regular nodes. During this phase of information collection, the adversary nodes may tamper with actual information collected by beacon node which leads towards the poor localization accuracy. This case can occur in any of node i.e., regular node or beacon node. The malicious regular node may transmit false data to the beacon node which may affect the Quality of Service (QoS). If the malicious node transmits redundant data, data aggregation scheme is applied in such cases. In order to maintain the credibility of regular sensor node and beacon node, a trust Evaluation Based model is proposed. In this model, a sensor node that has the trust value which is highest is considered as a reliable node and node with lowest trust value is discarded in the current round of simulation.

Considering that an unknown regular node N transmits location request signal $'Loc\_req'$ and a beacon node which is in the range of this node, transmits an acknowledge packet $'Loc\_ack'$ to the corresponding node. Here, a trust computation model for beacon node and regular node, vice-versa has been incorporated. At this stage, the values for direct and indirect trust are computed to obtain the comprehensive trust values. The value for the comprehensive trust can be expressed as:

$$C = \alpha D + \beta I \qquad (2)$$

Where C is the Comprehensive Trust value, D represents the value of Direct Trust (DT) and α is a weight factor for DT (Direct Trust). I denote the

Indirect Trust (IT) value, β is the weight factor for IT (Indirect Trust). In order to compute the DT of any node which consists of n attributes. A set of these attributes is denoted as $A = \{A_1, A_2, .A_i, .., A_n\}\,(0 \le A_i \le 1)$. Here, A is an attribute set containing various attributes. Each attribute value has specific impact on trust computation hence a weigh vector is defined as. Along with this, consider a time decay function $T(f)$ as follows:

$$T(f) = \begin{cases} 0 & k = 0, \\ 1 & k = 1, \\ T(f-1) - \left(\frac{1}{2}\right)^{f-1} & 1 \le f \le n \end{cases} \qquad (3)$$

Based on this, the direct trust value can be given as:

$$W_k^{(\mathcal{A},v)} = T(f) \times \sum_{i=1}^{n} v_i \times \mathcal{A}_i \qquad (4)$$

For computing the final trust value, combine all previous trust values. The final Direct Trust (DT) value can be computed as:

$$D_{xy} = \prod_{f=1}^{n} W_f^{(\mathcal{A},v)} \qquad (5)$$

Where D_{xy} signifies the Direct Trust value between source node x and destination node y.

Similarly, compute the indirect trust value for the network scenario as defined in Fig. 1. Below. In the scenario there is a source node, mediator node and a target node.

The source node, mediator node and target node are represented as S , M and τ respectively, I and D denotes the trust values of intermediate and direct trust. Consider three nodes which are anchor nodes that are very close to a node which is target node will be preferably considered as recommended node and are termed as R_l , R_m and R_n .
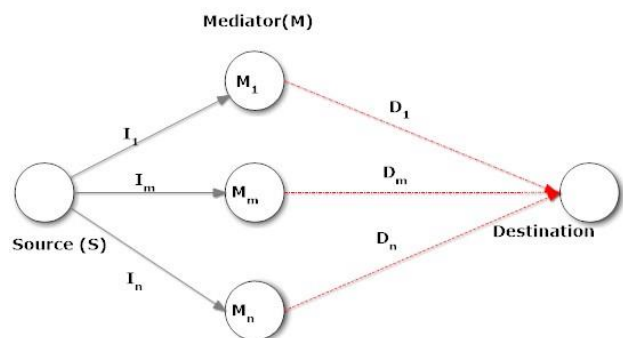

Figure. 1 Trust computation scenario

The most suitable node can be obtained based on trust value that is defined as:

$$I_{xy}(f,s,n) = \begin{cases} \frac{1}{2} & i = 1, \\ \frac{1+\sum_{i=1}^{n} T(f) \times s_i}{2+n} & i > 1 \end{cases} \quad (6)$$

Here n characterizes the overall nodes participated in the trust computation and s denotes the Boolean value which indicates that whether a node is considered for direct trust computation. Finally, the trust value for mediator node is expressed as:

$$M_{xy}(f,s,n,\mathcal{A},v) = \frac{1}{2}(\Phi + \Omega + \Psi) \quad (7)$$

Where $\Phi = (I_{xu}(f,s,n) \times D_{uy}(\mathcal{A},v))$, $\Omega = I_{xv}(f,s,n) \times D_{vy}(\mathcal{A},v)$, and $\Psi = I_{xw}(f,s,n) \times D_{wy}(\mathcal{A},v)$.

To calculate the final trust value, defining the information entropy of direct trust corresponding to their attributes, which is expressed as:

$$H_{sdd} = -D_{xy}(\mathcal{A},v) \times \log_2\left(D_{xy}(\mathcal{A},v)\right) - \left(1 - D_{xy}(\mathcal{A},v)\right) \times \log_2\left(1 - D_{xy}(\mathcal{A},v)\right) \quad (8)$$

Similarly, the information entropy of intermediate trust is measured and computed as:

$$\begin{aligned} H_{smm} = &-\mathcal{M}_{xy}(f,s,n,\mathcal{A},v) \\ &\times \log_2\left(\mathcal{M}_{xy}(f,s,n,\mathcal{A},v)\right) \\ &-\left(1 - \mathcal{M}_{xy}(f,s,n,\mathcal{A},v)\right) \\ &\times \log_2\left(1 - \mathcal{M}_{xy}(f,s,n,\mathcal{A},v)\right) \end{aligned} \quad (9)$$

With the help of these values of direct and intermediate node trust values, the weigh distribution values are obtained as follows:

$$\alpha = \frac{H_{smm}}{H_{smm} + H_{sdd}}$$

$$\beta = \frac{H_{sdd}}{H_{smm} + H_{sdd}} \quad (10)$$

These values can be replaced in above given Eq. (2) to obtain the final trust value. Based on these trust values, validation of node is done to check whether it can be considered for further process. There is a difference between actual distance and measured distance in Wireless Sensor Networks that follows the normal distribution, thus, the error function can be expressed as:

$$\sigma_N(ds) = ae^{-\frac{(ds - ds_0)^2}{b^2}} \quad (11)$$

$\sigma_N(ds)$ denotes the Gaussian function for distance ds. If value of ds and ds_0 are equal measurement error reaches to the maximum value. For simplicity, the distance measurement model for A_1 attribute which is expressed as:

$$p_1(ds) = \begin{cases} \frac{ds}{ds_0} & ds \leq ds_0 \\ \frac{ds - ds_0}{d_0} & ds > ds_0 \end{cases} \quad (12)$$

Defining the location reference of unknown node as

$$R = \left\{\left(j_1, k_1, d_1\right), \left(j_2, k_2, d_2\right), \ldots \left(j_i, k_i, d_i\right), \ldots \left(j_n, k_n, d_n\right)\right\}$$

here $\left(j_i, k_i\right)$ are the anchor node coordinates and d_i signifies the distance between and unknown node and anchor nodes. Here, the coordinates j and k are obtained by using least square and least square regression model as follows:

$$d_i^2 = (j - j_i)^2 + (k - k_i)^2 + \eta$$

$$(j,k) = \arg\min \sum_{i=1}^{n}\left(\sqrt{(j_i - j_0)^2 + (k_i - k_0)^2} - d_i\right)^2 \quad (13)$$

Where η denotes the measurement error as $\eta \square U\left(-\square\right)$ where $\cdot$ is the maximum error of measurement given as:

$$max\left|d_i - \sqrt{(j - j_i)^2 + (k_i - k)^2}\right| \leq \epsilon \quad (14)$$

Now, RSSI (Received Signal Strength Indicator) signal attenuation model and the distance ratio between transmitter and receiver can be defined as:

$$\frac{d_r^i}{d_r^j} = 10^{(RSSI(d_r^i) - RSSI(d_r^j))/10\lambda} \quad (15)$$

The coordinates of three anchor nodes are given as $\left(c_1, d_1\right), \left(c_2, d_2\right)$ and $\left(c_3, d_3\right)$ unknown node coordinate is $\left(c, d\right)$, corresponding distance of anchor node is obtained as dd_1, dd_2 and dd_3 as:

$$d_1 = \sqrt{(c - c_1)^2 + (d - d_1)^2}$$
$$d_2 = \sqrt{(c - c_2)^2 + (d - d_2)^2}$$
$$d_3 = \sqrt{(c - c_3)^2 + (d - d_3)^2} \quad (16)$$

With the help of least square based method, the coordinates of anonymous node can be obtained as:

$$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 2(c_1 - c_3)2(d_1 - d_3) \\ 2(c_2 - c_3)2(d_2 - d_3) \end{bmatrix}^{-1}$$
$$\begin{bmatrix} c_1^2 - c_3^2 + d_1^2 - d_3^2 + dd_3^2 - dd_1^2 \\ c_2^2 - c_3^2 + d_2^2 - d_3^2 + dd_3^2 - dd_2^2 \end{bmatrix} \quad (17)$$

3.1.3. Clustering and cluster head selection

The cluster formation and Cluster Head Selection model are presented here. First the energy depletion model for source and sink/next-hop to perform specific task based on the size of data packet is presented. Consider both free space model and multi-path fading channels based on the sources and sink module distances. According to the network density, consider a distance threshold as d_0, if the valued of distance between source and sink/next-hop is smaller than the given threshold distance then consider a model which is free-space (f_s) or else use multipath model. Let us consider that electronic circuit and amplifier use the energy E_{elec} in multipath (mp) model and free space $\bullet_{fs}$ respectively. Thus, on this assumption, the energy which is required by radio to communicate the q bit data over a distance s is expressed as:

$$E_t(l,d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2, \; for \; d < d_0 \\ lE_{elec} + l\epsilon_{mp}d^4, \; for \; d \geq d_0 \end{cases}$$
$$(18)$$

Similarly, the radio would consume energy while receiving q bit data over a given distance s and expressed as:

$$E_R(l) = lE_{elec} \quad (19)$$

The energy utilization parameter E_{elec} relies on several parameters such as modulation scheme, encoding model, filtering and signal spreading, similarly, the amplifier energy is given as $\dfrac{\bullet_{fs}s^2}{\bullet_{mp}s^4}$ depends on the transmitter and receiver distance. In the heterogeneous network considerations, the preliminary energy is arbitrarily assigned to each as $\left[E_0, E_0(q + \alpha_{max}) \right]$ where E_0 is the lower bound and α_{max} decide the maximum energy value. The total energy of heterogeneous network is given as:

$$E_{total} = \sum_{i=1}^{N} E_0(1 + \alpha_i) \quad (20)$$

Table 2. Initial packet structure

Node identification	Residual energy	Distance from BS
ID_i	$E_{res}(i)$	$d(ch_i, bs)$

Assuming that each non-cluster head node transmit the L bit size data packet in each round. Hence, the total dissipated energy for each round can be expressed as:

$$E_{round} = L\big(2NE_{elec} + kE_{mp}d_{toBS}^4 + NE_{DA} + NE_{fs}d_{toCH}^2\big) \quad (21)$$

By means of this, the number of cluster heads which is optimal can be obtained as:

$$K_{opt} = \frac{M}{d_{toBS}^2} \frac{\sqrt{N}}{\sqrt{2\pi}} \frac{\sqrt{E_{fs}}}{\sqrt{E_{mp}}} \quad (22)$$

Based on this process, the node broadcasts the JOINT request to the neighbouring node which will be elected as Cluster Head. The neighbouring nodes/participating node receives this message from the neighbouring nodes. The best criteria for selecting the cluster head are the cluster head stability residual energy and distance ranges from the next cluster head or base station. Here, the clusters which are nearest to the base station have the size which is smaller because of less energy requirement and less distance to transmit the packet. In this work, a model has been introduced which can be used to assign the cluster radius which is expressed as:

$$R_{opt} = \left(1 - \frac{d_{max} - d(i,BS)}{d_{max} - d_{min}} \right) \times r_0 \times \sqrt{\frac{E_{residual}}{E_0}} \quad (23)$$

The Initial packet structure for communication model is shown in Table 2.

The communication model of proposed approach is as follows:

Measure the distance between the selected cluster head and base station as $d(ch_i, bs)$ and use multi-hop routing if the distance is measured as greater than the Threshold d_0. Next if the value of $d(ch_i, bs)$ is less comparatively lesser than the threshold value, then Cluster Head is eligible to communicate directly to the base station.

3.1.4. Data aggregation

Data aggregation scheme plays important role in WSN which helps to enhance the network lifetime and QoS(Quality of Service). Assign a variable ψ as the successful packet delivery rate to the next hop

which is used to maintain the QoS (Quality of Service) level. Let us consider that transmission of γ packets to the next hop after successful localization and trust evaluation, the transmission reliability is denoted as $\psi_i = 1 - (1 - p_{ij})^\gamma$. With the help of this, the network transmission reliability can be considered as $\prod_k \left(1 - (1 - p_{ij})^\gamma\right)$. Here, the focus is on improving the network lifetime while maintaining the transmission reliability, hence, the energy utilization minimization objective function is defined as:

$$\begin{cases} Max(T) = Min\left[\max_{0<i<N}(E_i)\right] \\ E_i = P_i^t E_t + P_i^r E_r \\ \frac{\sum_k \sum_i \sigma(v_n^i)}{P} \geq \epsilon \end{cases} \quad (24)$$

p_i^r denotes the packet size for receiving node and p_i^t denotes the packet size for transmitting node, E_t is energy utilized by transmitter node, E_r is the energy utilized by receiver node and overall energy consumption for node i is given as E_i $\sigma(v_n^i)$ is the indicator successful packet delivery. For successful delivery $\sigma(v_n^i) = 1$ and $\sigma(v_n^i) = 0$ if packets are not delivered. In order to meet the aggregation criteria, there is a usage of "average" aggregation function. Let us consider that a time schedule for all nodes in the cluster is given as $\{\tau_1, \tau_2, ..., \tau_k\}$. Assuming that the current cluster consist of k number of nodes and the cluster head of current cluster aggregates the data for the considered time period T as $X = \{X_1, X_2, ..., X_n\}$. The average Function based aggregation can be expressed as:

$$P_{avg} = Avg\left(\sum_{j=1}^{n} X_j\right) \quad (25)$$

3.2 Results and discussion

The experimental analysis is described using proposed hybrid approach of secure localization, data aggregation, and energy aware cluster head selection. Hence the proposed hybrid method is executed using a simulation tool such as MATLAB which is installed on windows operating systems. The windows platform has the following specifications like 8 Giga Byte of RAM, Intel Core i3 Processor and 1Terra Byte of hard drive. The simulation parameters are described below in the given Table 3.

Consider a network scenario where 20 nodes are considered as anchor node and the performance is

Table 3. Parameters for simulation

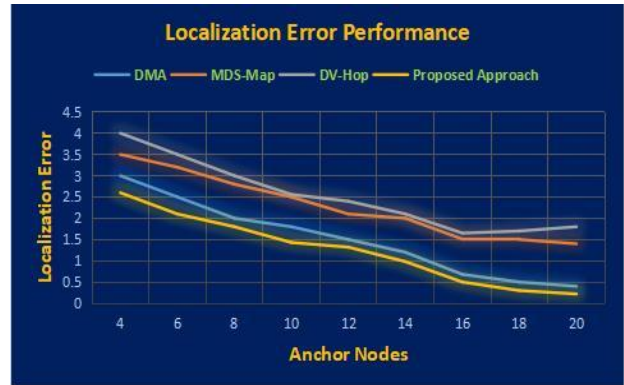| Simulation Parameter	Considered Parameter
Network area	100 m
Number of nodes	100-300
Initial energy	1.5J
Transmitter/ receiver power	50nJ/bit
Number of anchor nodes	20
e_{mp}	0.0013 pJ/bit/m²
ε_{fs}	10 pJ/bit/m²

Figure. 2 Localization error performance

measured considering the factors of localization error and energy consumption. The obtained performance is put into comparison with existing techniques such as DV-Hop(Distance Vector-Hop), MDS-Map(Multidimensional Scaling) and DMA(Distance Mapping Algorithm) as mentioned in [36]. Below given Fig. 2 shows the comparative analysis of localization performance.

Above given Fig. 2 shows that as the anchor nodes are increasing the localization error decreases. The existing techniques have more error when compared to proposed approach because proposed approach uses correlation between sensor node distances. Moreover, the security module helps to consider only the reliable information hence false location is avoided in this model. The experimental analysis depicts the average localization error is obtained as 1.51, 2.28, 2.52 and 1.25 using DMA (Distance Mapping Algorithm), MDS-Map (Multidimensional Scaling), DV-Hop (Distance Vector-Hop) and Proposed Approach, respectively. The inaccurate localization affects several performance parameters of the network.

Similarly, the energy depletion performance for data aggregation is measured. Fig. 3. given below depicts the energy dissipation performance for 20 anchor node scenarios. Similar to previous experiment, the performance is compared with existing techniques as described in [36].
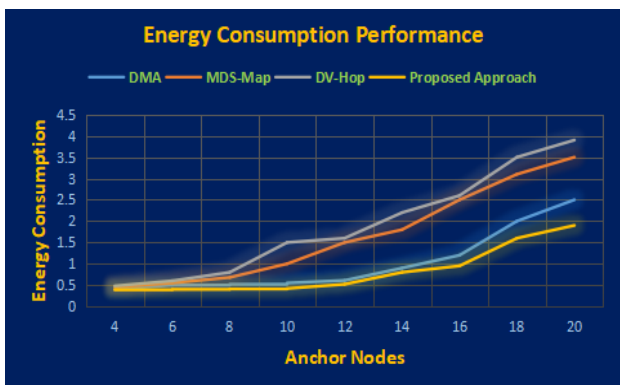
Figure. 3 Energy consumption performance for localization

Figure. 4 Alive node performance for 100 nodes

Figure. 5 Alive node performance for 300 nodes

As per this experiment, the increased number of anchor nodes leads to increase the energy dissipation because all anchor nodes transmit the packet to the nearby nodes to estimate the location. The proposed approach achieves better performance while considering the secure localization whereas existing approaches don't consider the security aspects. This analysis shows that without considering security mode, better functioning can be achieved by diminishing the energy depletion. However, still the proposed approach achieves desired performance while maintaining security and efficient localization. The average energy consumption is obtained as 1.02, 1.67, 1.90 and 0.82 using DMA (Distance Mapping Algorithm), MDS-Map (Multidimensional Scaling), DV-Hop (Distance Vector-Hop) and Proposed Hybrid Approach, respectively.

Further, the performance for inter-cluster and intra-cluster communication is measured. The data which is aggregated is further transferred to the corresponding base station. Thus, here in this process the complete network consumes energy and for multiple rounds of simulation the nodes gradually start dying because of not enough energy to transmit the data packet. Comparing the obtained performance with existing techniques such as LEACH (Low-energy adaptive clustering hierarchy), O-LEAC H(Optimization Low Energy Adaptive Clustering Hierarchy), Bee Cluster, and OK-Means (Optimal K-means) as mentioned in [37].

Fig. 4 shows the comparative analysis in terms of alive nodes for 100 nodes. The simulation is performed for 5000 rounds and measured total alive nodes after finishing the simulation. Initially, there is less communication and each sensor node has efficient energy hence for 500 nodes number of dead nodes are not reported. During this simulation, the increase in simulation round leads to the energy depletion which causes rise in dead node count. This experiment shows that average number of alive nodes
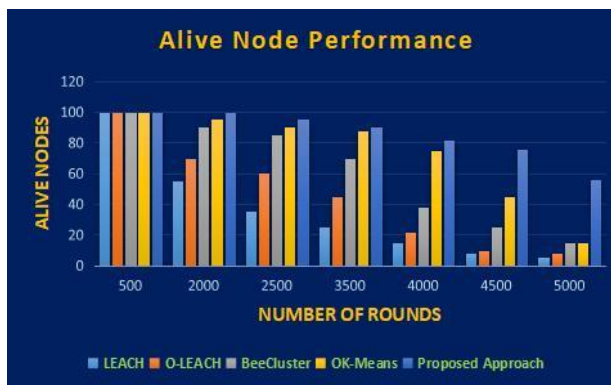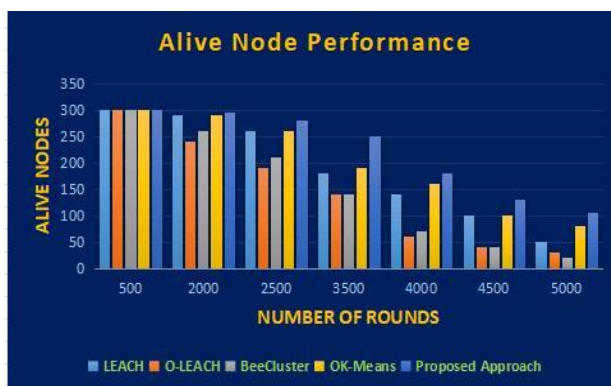
are obtained as 35, 45, 60, 73, and 56 using LEACH (Low-energy adaptive clustering hierarchy), O-LEAC H(Optimization Low Energy Adaptive Clustering Hierarchy), Bee Cluster, and OK-Means (Optimal K-means) and Proposed Hybrid Approach, respectively. The performance for 300 node scenarios is also measured and depicted in Fig. 5.

In this experiment, consider 300 nodes scenario for 5000 rounds, the average alive nodes are obtained as 188, 142, 148, 197 and 220. Further, comparing the performance with reference to energy consumption for varied number of node scenario as depicted in Fig. 6.
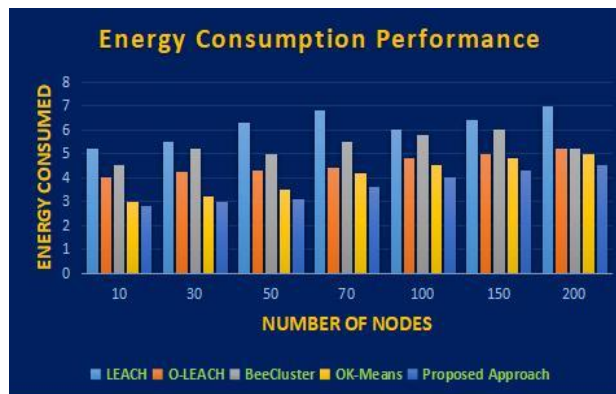

Figure. 6 Performance on energy consumption

The average energy consumption as 6.17, 4.56, 5.31, 4.02 and 3.61 using LEACH (Low-energy adaptive clustering hierarchy), O-LEACH (Optimization Low Energy Adaptive Clustering Hierarchy), Bee Cluster, and OK-Means (Optimal K-means) and Proposed Approach has been achieved in this experiment respectively. The proposed methodology uses a combination of node localization, data aggregation and energy-aware cluster formation which helps to reduce the overall energy consumption.

4. Conclusions

In this approach, the focus is on improving the network performance with the reference to Network Lifetime, Minimizing the energy dissipation and incorporating the security. This work is divided into three stages where energy aware Sensor Node localization model is presented and in the next phase Cluster Formation and Cluster Head selection is achieved. Finally, data aggregation scheme is presented to discard the data redundancy. The localization model uses RSSI (Received Signal Strength Indicator) based scheme and trust evaluation model, the clustering scheme uses energy related parameters of sensor nodes and aggregation is performed using "average" function. The comparative analysis demonstrates the significance of proposed approach due to its quality to improve the performance of the network by decreasing the energy depletion and increasing the lifespan of a network. In future, the lightweight cryptography schemes can be incorporated to improve the data security.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, methodology, original draft preparation and simulation is carried out by first author. Supervision, guidance and review has been carried out by the second author.

Acknowledgments

References

[1] P. Donta, A. Tarachand, and C. Sekhar, "Machine learning algorithms for wireless sensor networks: A survey", *Information Fusion*, Vol. 49, pp. 1-25, 2019.

[2] H. Yetgin, K. Cheung, M. E. Hajjar, and L. Hanzo, "A survey of network lifetime maximization techniques in wireless sensor networks", *IEEE Communications Surveys & Tutorials*, Vol. 19, No. 2, pp. 828-854, 2017.

[3] H. Kobo, A. A. Mahfouz, and G. Hancke, "A survey on software-defined wireless sensor networks: Challenges and design requirements", *IEEE Access*, Vol. 5, pp. 1872-1899, 2017.

[4] S. Randhawa and S. Jain, "Data aggregation in wireless sensor networks: Previous research, current status and future directions", *Wireless Personal Communications*, Vol. 97, No. 3, pp. 1-71, 2017.

[5] S. Wan, Y. Zhang, and J. Chen, "On the Construction of Data Aggregation Tree with Maximizing Lifetime in Large-Scale Wireless Sensor Networks", *IEEE Sensors Journal*, Vol. 16, pp. 1-1, 2016.

[6] T. Kalidoss, K. Kanagasabai, R. Logambigai, S. Munuswamy, S. Ganapathy, and K. Arputharaj, "Energy Aware Cluster and Neuro-Fuzzy Based Routing Algorithm for Wireless Sensor Networks in IoT", *Computer Networks*, Vol. 151, pp. 211-223, 2019.

[7] A. S Rostami, M. Badkoobe, F. Mohanna, H. Keshavarz, A. A. R Hosseinabadi, and A. Sangaiah, "Survey on clustering in heterogeneous and homogeneous wireless sensor networks", *The Journal of Supercomputing*, Vol. 74, pp. 277-323, 2018.

[8] D. Sharma and A. Bhondekar, "Traffic and energy aware routing for heterogeneous wireless sensor networks", *IEEE Communications Letters*, Vol. 22, No. 8, pp. 1608-1611, 2018.

[9] H. Mostafaei, "Energy-efficient algorithm for reliable routing of wireless sensor networks", *IEEE Transactions on Industrial Electronics*, Vol. 66, No. 7, pp. 5567-5575, 2019.

[10] G. Han, J. Jiang, C. Zhang, T. Duong, M. Guizani, and G. Karagiannidis, "A survey on mobile anchor node assisted localization in wireless sensor networks", *IEEE Communications Surveys & Tutorials*, Vol. 18, No. 3, pp. 2220-2243, 2016.

[11] J. Kuriakose, S. Joshi, R. Raju, and A. Kilaru, "A review on localization in wireless sensor networks", *In Advances in Signal Processing and Intelligent Recognition Systems*, Vol. 264, pp. 599-610, 2014.

[12] A. Belay, H. Lin, G. Tarekegn, and S. S. Jeng, "Applying deep neural network (DNN) for robust indoor localization in multi-building

environment", *Applied Sciences*, Vol. 8, No. 7, p. 1062, 2018.

[13] T. Chowdhury, C. Elkin, V. Devabhaktuni, B. D Rawat, and J. Oluoch, "Advances on localization techniques for wireless sensor networks: A survey", *Computer Networks*, Vol. 110, pp. 284-305, 2016.

[14] A. Paul and T. Sato, "Localization in wireless sensor networks: A survey on algorithms, measurement techniques, applications and challenges", *Journal of Sensor and Actuator Networks*, Vol. 6, No. 4, p. 24, 2017.

[15] B. Jan, H. Farman, H. J. Huma, B. Montrucchio, and S. Ali, "Energy efficient hierarchical clustering approaches in wireless sensor networks: A survey", *Wireless Communications and Mobile Computing*, Vol. 2017, pp. 1-14, 2017.

[16] A. Coluccia and A. Fascista, "Hybrid TOA/RSS Range-Based Localization with Self-Calibration in Asynchronous Wireless Networks", *Journal of Sensor and Actuator Networks*, Vol. 8, No. 2, p. 31, 2019.

[17] S. Pandey and S. Varma, "A range-based localization system in multihop wireless sensor networks: a distributed cooperative approach", *Wireless Personal Communications*, Vol. 86, No. 2, pp. 615-634, 2015.

[18] O. Cheikhrouhou, G. Bhatti, and R. Alroobaea, "A hybrid DV-hop algorithm using RSSI for localization in large-scale wireless sensor networks", *Sensors*, Vol. 18, No. 5, p. 1469, 2018.

[19] T. Le, L. C. Tran, X. Huang, C. Ritz, E. Dutkiewicz, S. Phung, A. Bouzerdoum, and D. Franklin, "Unbalanced Hybrid AOA/RSSI Localization for Simplified Wireless Sensor Networks", *Sensors*, Vol. 20, No. 14, p. 3838, 2020.

[20] L. Zhang, Z. Yang, S. Zhang, and H. Yang, "Three-dimensional localization algorithm of WSN nodes based on RSSI-TOA and single mobile anchor node", *Journal of Electrical and Computer Engineering*, Vol. 2019, pp. 1-8, 2019.

[21] Y. Lv, W. Liu, Z. Wang, and Z. Zhang, "WSN Localization Technology Based on Hybrid GA-PSO-BP Algorithm for Indoor Three-Dimensional Space", *Wireless Personal Communications*, Vol. 114, 2020.

[22] P. Singh and N. Mittal, "An Efficient Localization Approach for WSNs using Hybrid DA-FA Algorithm", *IET Communications*, Vol. 14, 2020.

[23] Y. Meng, Q. Zhi, Q. Zhang, and N. Yao, "A Two-Stage Particle Swarm Optimization Algorithm for Wireless Sensor Nodes Localization in Concave Regions", *Information*, Vol. 11, No. 10, p. 488, 2020.

[24] B. Mukhopadhyay, S. Srirangarajan, and S. Kar, "Robust range-based secure localization in wireless sensor networks", In: *Proc. of GLOBECOM*, pp. 1-6, 2019.

[25] B. Kang, P. Nguyen, V. Zalyubovskiy, and H. Choo, "A distributed delay-efficient data aggregation scheduling for duty-cycled WSNs", *IEEE Sensors Journal*, Vol. 17, No. 11, pp. 3422-3437, 2017.

[26] W. Fang, X. Wen, J. Xu, and J. Zhu, "CSDA: a novel cluster-based secure data aggregation scheme for WSNs", *Cluster Computing*, Vol. 22, No. 3, pp. 5233-5244, 2019.

[27] J. Zhang, P. Hu, F. Xie, J. Long, and A. He, "An energy efficient and reliable in-network data aggregation scheme for WSN", *IEEE Access*, Vol. 6, pp. 71857-71870, 2018.

[28] W. Alghamdi, M. Rezvani, H. Wu, and S. Kanhere, "Routing-aware and malicious node detection in a concealed data aggregation for WSNs", *ACM Transactions on Sensor Networks*, Vol. 15, No. 2, pp. 1-20, 2020.

[29] I. Mosavvar and A. Ghaffari, "Data aggregation in wireless sensor networks using firefly algorithm", *Wireless Personal Communications*, Vol. 104, No. 1, pp. 307-324, 2019.

[30] H. Song, S. Sui, Q. Han, H. Zhang, and Z. Yang, "Autoregressive integrated moving average model–based secure data aggregation for wireless sensor networks", *International Journal of Distributed Sensor Networks*, Vol. 16, No. 3, p. 155014772091295, 2020.

[31] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand, and A. H. Gandomi, "Residual energy-based cluster-head selection in WSNs for IoT application", *IEEE Internet of Things Journal*, Vol. 6, No. 3, pp. 5132-5139, 2019.

[32] K. N. Dattatraya and K. R. Rao, "Hybrid based cluster head selection for maximizing network lifetime and energy efficiency in WSN", *Journal of King Saud University-Computer and Information Sciences*, 2019.

[33] H. Robinson, G. Julie, and L. Son, "Probability-based cluster head selection and fuzzy multipath routing for prolonging lifetime of wireless sensor networks", *Peer-to-Peer Networking and Applications*, Vol. 12, No. 5, pp. 1061-1075, 2019.

[34] S. Murugaanandam and V. Ganapathy, "Reliability-based cluster head selection methodology using fuzzy logic for performance

improvement in WSNs", *IEEE Access*, Vol. 7, pp. 87357-87368, 2019.

[35] A. Baradaran and K. Navi, "HQCA-WSN: High-quality clustering algorithm and optimal cluster head selection using fuzzy logic in wireless sensor networks", *Fuzzy Sets and Systems*, Vol. 389, pp. 114-144, 2019.

[36] R. Tan, Y. Li, Y. Shao, and W. Si, "Distance mapping algorithm for sensor node localization in WSNs", *International Journal of Wireless Information Networks*, Vol. 27, No. 2, pp. 261-270, 2019.

[37] S. E. Khediri, W. Fakhet, T. Moulahi, R. Khan, A. Thaljaoui, and A. Kachouri, "Improved node localization using K-means clustering for Wireless Sensor Networks", *Computer Science Review*, Vol. 37, p. 100284, 2020.