MIPSOE – Markov Integrated PSO Encryption Algorithm for Secure Data Aggregation

Reshma S[a], Shaila K[b], Thippeswamy B M[c], Venugopal K R[d]

[a]Research Scholar, VTU-RRC, Belagavi, Karnataka, India
bProfessor, Department of Electronics and Communication Engineering, Vivekananda Institute of Technology, Bangalore, Karnataka, India
[c]Professor, School of Electrical Engineering and Computing, Science and Technology University, Adama, Ethiopia
[d]Vice Chancellor, Bangalore University, Bengaluru, Karnataka, India

ABSTRACT:Various clustering algorithm exists in Wireless Sensor Networks concerned on balancing energy utilization. Many research issues deviate towards the formation of clusters based on energy, distance, and another sensor node's resource parameters. In this article, the proposed protocol is composed of two phases. In the first phase, clusters are formed based on Particle Swarm Optimization and Markov's Random Field mathematical calculation. The second phase generates a key, where the secret key is used for encryption technique. The proposed protocol is implemented in the NS2 simulator. When comparing the existing protocol with the proposed MIPSOE protocol it is inferred that there is an improvement in terms of network lifetime, throughput, delay, and packet delivery ratio.

Keywords: Markov Random Field, Particles Swarm Optimization, Wireless Sensor Networks

1. Introduction

The sensor nodes are self-configured and are connected to the internet for communication, which is referred to as the Internet of Things. The Internet of Things is a part of Wireless Sensor Networks. Wireless Sensor Networks are open and unprotected communication channels between sensor nodes. The network is forced to several intruder attacks and also unprotected network sensor nodes are vulnerable to limited battery usage. Therefore, it is necessary to balance energy usage in the network without network disconnectivity.

The existing algorithms [1] LD^2 FA-PSO designed a lightweight scheme to mitigate Black-hole attack [14], [2] mitigates multi-layer flooding attack and minimizes energy utilization by monitoring residual energy status and [3] detects Sybil attack in large WSNs and reduces false alarm rate. Reduced network clustering and balanced consumption clustering is a process of connecting sensor nodes into one group with one cluster head. There are different algorithms to form a cluster and to elect cluster heads such as LEACH [4], Particle Swarm Optimization [5], GSA [6], and MOEA [7].

These algorithms are a trade-off between energy, latency, and data protection. So, it is necessary to develop a mechanism that should be capable of monitoring energy, delay, and data protection. Therefore, in our work, a protocol is designed in such as to fulfil the constraints of WSNs. The protocol forms cluster with Integrating Particle Swarm Optimization and Markov Random Field concept and protects data with enhanced encryption technique which generates a key based on cluster and energy density.

The Particle Swarm Optimization algorithm searches for the optimal best path based on swarm intelligence. Initially, obtains the basic information about the sensor nodes and the environment and applies fitness function. The global best fitness value is considered to find the similarity of the sensor nodes and forms the clusters using the cosine similarity function. This PSO algorithm is incorporated with Markov Rando Field to obtain the cluster head in the proposed protocol.

The rest of the article is organized as, Section 2 discusses works related to secure data aggregation and background work is presented in Section 3. The MIPSOE System Flow diagram is depicted and explained in Section 4. Simulation results and comparison with the existing protocol are analyzed in Section 5. Finally, Section 6 concludes the research work.

2. Related Work

This section discusses existing protocols for secure data transmission in Wireless Sensor Networks.
Zahhad et al., [8] proposed Mobile Sink based adaptive Immune Energy-Efficient Clustering protocol. This protocol divides a network into several zones and guides CH's to minimize energy utilization and packet overhead. This protocol obtains a minimal number of Cluster Heads and increases the stability of the network. Krishnan et al., [9] implemented an efficient clustering algorithm. This algorithm surveyed the static sink and its disadvantage. To overcome static sink functionality, the algorithm introduces multiple sinks. Thus, it reduces the number of hop transmission in the network. So, it reduces energy consumption and delays for data transmission. This algorithm elects CH based on nodes that possess higher energy, thus leading network disconnectivity.

Bala et al., [10] presented a multi-objective meta-heuristic approach for energy-efficient secure data aggregation which involves three phases. In the first phase, the clusters are formed. In the next phase, the protocol selects a secure node and finally, the data is aggregated in the third phase. The clusters are formed uniformly and it is reformed based on resource utilization. The BS distributes a public key to all the sensor nodes and CH broadcasts a private key to its respective cluster members. The CH is responsible for validating a private key of sensor nodes during data transmission. But, this protocol is not focusing on delay during data transmission.

Zhang et al., [11] preferred a multicast routing on a software-defined network and multicast routing is obtained by virtualization of the network functionality. Multicast algorithms are structured for a static and dynamic environment. Multicast algorithms are developed by considering the cases of unchanged and changed multicast bandwidth cases. Algorithms are framed by utilizing the Network Function Virtualization. The required network apparatus has been designed by the software mechanism. Implementation of the work is done on Virtual Network Overlay (ViNO) carried on Smart Applications on Virtual Infrastructures (SAVI) testbed. The optimization of multicast routing is achieved by emphasizing on joint scheduling of multicast assemblies. Multicast routing topology is verified for cost and time analysis parameters. Setting the priority level for each routing path in multicast topology becomes a challenging task in the above multicast routing.

Chuan et al., [12] proposed an algorithm for industrial WSNs whose sink is movable. To overcome the hotspot challenge for heterogeneous network, this algorithm defines tree-based topology for data collection. Data gathering in a heterogeneous environment is achieved by assuming rendezvous and sub rendezvous points in the network architecture. Rendezvous nodes are root nodes of the tree topology network and some particular nodes based on the weights are defined as sub rendezvous points. These points act as a data collection entity and are successful in reducing the network's energy consumption. The responsibility of these points is also shared by other members to overcome the hotspot problem of these points. This algorithm is successful in achieving the energy efficiency of the heterogeneous network and finds more suitable in different networks of industrial applications.

3. Background

Khalid et al., [13] presented LSDAR protocol to optimize the routing path with lesser energy utilization. With the help of the LSDAR protocol, a large sparse network is broken into clusters with a random radius. The malicious node intrusion is minimized and provides security using a One-Time Pad (OTP) encryption method. The drawback of LSDAR protocol is unbalanced cluster formation and LSDAR protocol involves varying cluster size in the network. In large clusters, data communication happens frequently and mitigates energy in an optimal path which leads to network disconnectivity. The proposed work focuses on balancing cluster size, energy utilization, delay, and network lifetime.

4. System and Mathematical Model

Figure 1 shows the MIPSOE system flow diagram. The MIPSOE protocol involves 3 major phases viz., Markov Integrated with PSO (MIPSO) Phase, Encryption Phase, and Optimal Path Computation Phase respectively. The MIPSOE Phase integrates Particle Swarm Optimization and Markov Random Field Calculation technique. The encryption phases encrypt generated data using XOR operation, key-value generated, and are stored in the base station routing table.

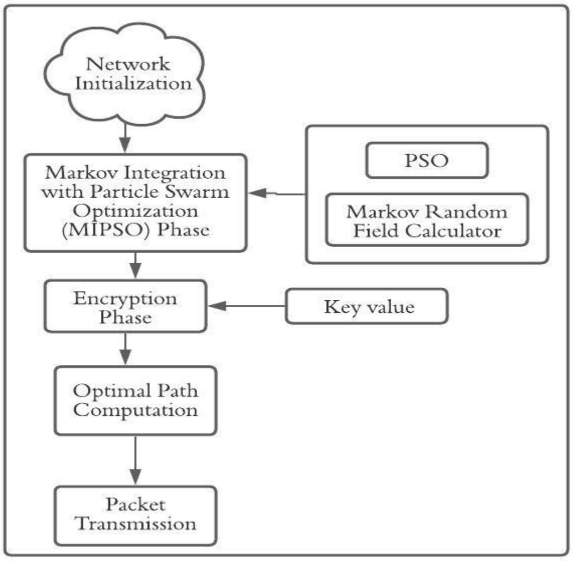
Network Initialization
Markov Integration with Particle Swarm Optimization (MIPSO) Phase
PSO
Markov Random Field Calculator
Encryption Phase
Key value
Optimal Path Computation
Packet Transmission

Fig. 1 MIPSOE System Flow Diagram

Phase 1: Markov Integrated with PSO (MIPSO)

The PSO algorithm involves a swarming nature to group the birds and to train the birds to follow the optimal route. This type of intelligence training is called swarm intelligence. In our work, the PSO algorithm is used to train the sensor node's intelligence and form the cluster based on the resource constraints of the sensor node. Markov Random Field receives optimal decision data of all neighbor nodes obtained by the PSO algorithm in the cluster and votes for the highest value to elect the CH.

The MIPSO algorithm works as follows:

Step 1 : Initialize solutions for sensor nodes where sensor nodes are considered as a particle.

Solution 1: Compute Residual Energy (R_E)

$$R_E = I_E - (E_T + E_R) \tag{1}$$

Where, R_E is residual energy, I_E is initial energy, E_T is transmission energy and E_R is receiving the energy of a sensor node.

Solution 2: Node Position (P)

$$P_i = (X_i, Y_i) \tag{2}$$

Where, P_i is a position of node i, X_i is the X coordinate of node i and Y_i is Y coordinate of node i.

Solution 3: Node Degree (N_D)

$$N_D = |\{\mathrm{Neighbor}_{\mathrm{nodes}}\}| \tag{3}$$

Where, N_D is a Node Degree, $\{\mathrm{Neighbor}_{\mathrm{nodes}}\}$is a set of Neighbor Nodes and $|\{\mathrm{Neighbor}_{\mathrm{nodes}}\}|$ is a count of Neighbor Nodes.

Solution 4: Distance (D)

$$D(\mathrm{node}(i), B_S) = \sqrt{\left(X_{\mathrm{node}(i)} - X_{B_S}\right)^2 + \left(Y_{\mathrm{node}(i)} - Y_{B_S}\right)^2} \tag{4}$$

where, $D(\mathrm{node}(i), B_S)$ is a distance from node i to base station, $X_{\mathrm{node}(i)} - X_{B_S}$ is a X co-ordinate of node i and base station and $Y_{\mathrm{node}(i)} - Y_{B_S}$is a Y co-ordinate of node i and base station.

Step 2: Apply Fitness function (f)

$$f_{(\mathrm{node}(i))} = \sigma_1 R_E + \sigma_2 D_{(\mathrm{node}(i),B_S)} + \sigma_3 N_D \tag{5}$$

Where, $f_{(\mathrm{node}(i))}$ is a fitness value of node i,σ_1, σ_2 and σ_3 are the weightage parameters.

Step 3: Use cosine similarity with the fitness value to form the clusters

$$(6) \qquad CS\left(f_{(\mathrm{node}(i))}, f_{(\mathrm{node}(i+1))}\right) = \frac{\sum_{i=1}^{m}\left(f_{(\mathrm{node}(i))} \cdot f_{(\mathrm{node}(i+1))}\right)}{\sqrt{\sum_{i=1}^{m} f_{(\mathrm{node}(i))}}^2 \sqrt{\sum_{i=1}^{m} f_{(\mathrm{node}(i+1))}}^2}$$

Step 4: Apply selection criteria query for fitness value of each node to obtain the probability of selection to elect Cluster Head (CH).

$$P(M_R) = \prod_{i=1}^{N} \emptyset(\mathrm{node}(i), \mathrm{node}(i+1)) \tag{7}$$

Where, $\emptyset$ is selected on the basis if the highest fitness value computed in step 2.

Step 5: Formulate Cluster Head (CH) set based on the probability

$$\mathrm{CH_S} = \frac{1}{\mathrm{Z}} \emptyset(\mathrm{node}(\mathrm{i}), \mathrm{node}(\mathrm{i}+1)) \tag{8}$$

Where, $\mathrm{CH_S}$ is a Cluster Headset, Z is a summation of $\mathrm{P}(\mathrm{M_R})$ obtained using equation 7 among the cluster.

Step 6: Elect Cluster Head (CH) based on Cluster Head (CH) set

$$\mathrm{G_i}(\mathrm{CH}) = \sigma \max_{0<\mathrm{i}<\mathrm{M}}(\mathrm{CH_s}(\mathrm{i})) \tag{9}$$

Where, $\mathrm{G_i}(\mathrm{CH})$ is selected CH for the Cluster $\mathrm{G_i}$and i varies from 1 to M clusters

Phase 2: Encryption

The base station generates a secret key and stores at the base station security routing table shown in Figure 2.

BS_node_id	Sensor_Node_Id	Cluster_ID	Secret_Key_Value

Fig. 2 Base Station Security Routing Table

The encryption algorithm is implemented by the following steps:

Step 1: Base Station shares generated secret key value with all the sensor nodes.
Step 2: Sensor nodes sense data and encrypts data with the secret key value using XOR operation and forward it to the BS.
Step 3: Base Station receives encrypted data and decrypts with the secret key value by retrieving it from the base station security routing table.

Phase 3: Optimal Path Computation

The MIPSO phase divides the network into clusters and elects the Cluster Head. So, a network is prepared for data transmission between sensor nodes and base stations. Whenever sensors sense data, the Optimal path computation phase eliminates data redundancy to provide aggregated data which helps to reduce energy utilization for redundant data transmission. This phase also encrypts data as discussed in Phase 2 (Encryption Phase) before transmitting data to CH. The CH forwards data to the base station and the base station in turn decrypts data as discussed in phase 2. Finally, data will be used as required in the application.

5.Simulation and Performance Evaluation

The NS2 simulator is used for implementing MIPSOE protocol and the simulation is done for various network dimensions from 50 to 300 nodes over an area of 1000m * 1000m. The MIPSOE protocol is compared with the LSDAR protocol. The simulation settings for the MIPSOE protocol are shown in Table I.

Table I. Simulation settings

Simulation Variables	Values
Set of Nodes	50, 100, 150 200,250,300
Network dimension	1000m*1000m
Node's Energy – Initial	1 J

Node's Energy utilized for Transmission	0.016 J
Node's Energy utilized for Reception	0.018 J
Simulation Duration	20000s

5.1 Performance Metric

a) **Energy Consumption:** The fair utilization of energy for data transmission in the networks.

b) **End-to-End Delay:** The latency for delivering data from the source node to the base station within a network.

c) **Packet Drop Ratio:** The number of packets dropped during data transmission due to an intruder enters inside a network.

5.2 Performance Evaluation

Figure 3 depicts the energy utilization of the MIPSOE protocol with the comparison of LSDAR protocol. Table II represents the MIPSOE numerical comparison results of Energy utilization with LSDAR protocol. The LSDAR protocol encrypts sensed data with random value on every hop and it leads to more computation. This computation overhead increases energy utilization and time complexity. In MIPSOE protocol BS generates a random value and is cached at its routing table. Therefore, it helps to encrypt once at the data originated sensor node and decrypted at base station with the support of routing table information. The graph shows linear changes of both protocols as node size increases but MIPSOE protocol uses less average energy in the network when compared with LSDAR protocol due to lesser computation. Hence, MIPSOE proves that an improvement of 5% energy utilization is achieved over LSDAR protocol.

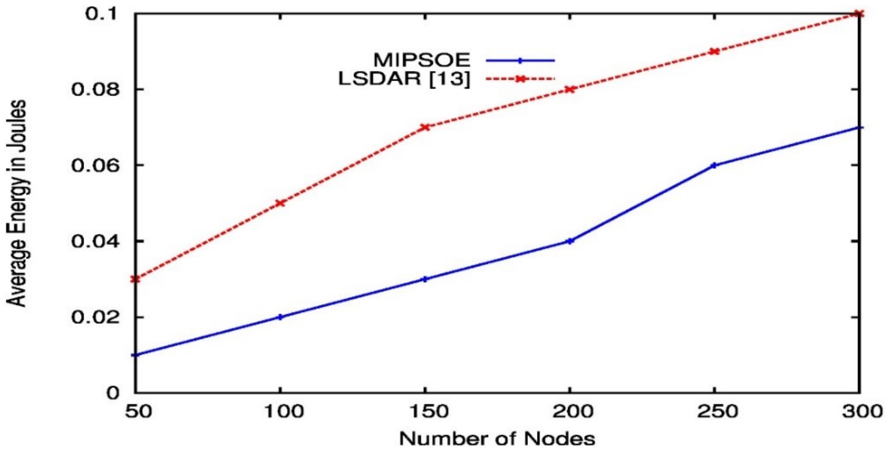

Fig.3 Energy Utilization of MIPSOE and LSDAR Protocol

Table II : Energy Utilization

Number of Nodes	Energy Utilization	
	LSDAR	MIPSOE
50	0.03	0.01
100	0.05	0.02
150	0.07	0.03
200	0.08	0.04
250	0.09	0.05
300	0.11	0.06

The End-to-End Delay of MIPSOE and LSDAR protocol during data transmission is depicted numerically and pictorially in Table III and Figure 4 respectively. Due to the balanced cluster size achieved using PSO and cosine similarity function. The energy is balanced optimally and CH is selected based on the Markov technique and highest PSO solution value. This mechanism helps to forward data through an optimal channel. Thus, helps to minimize an End-to-End Delay. Therefore, MIPSOE exhibits 8% End-to-End Delay compared with LSDAR protocol.

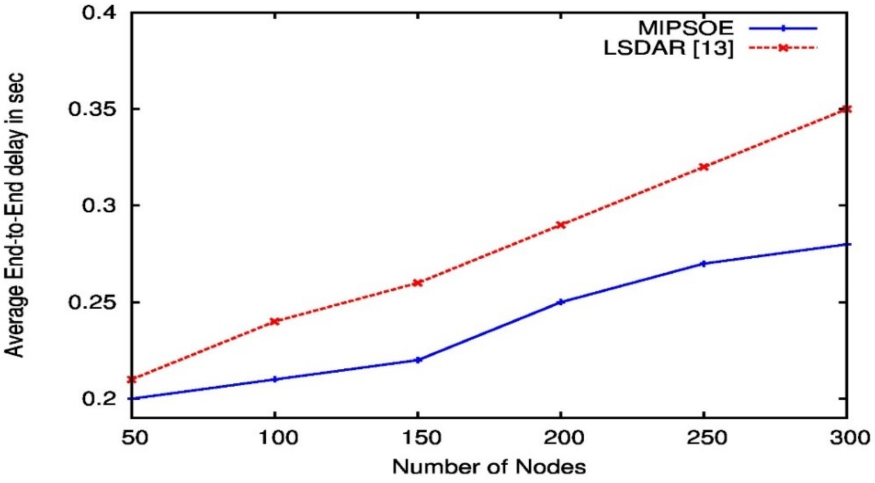

Fig. 4 Average End-to-End Delay of MIPSOE and LSDAR Protocol Vs. Different Network Size

Table III Average End-to-End Delay

Number of Nodes	Average End-to-End Delay	
	LSDAR	MIPSOE
50	0.21	0.20
100	0.24	0.21
150	0.26	0.22
200	0.29	0.25
250	0.32	0.27
300	0.35	0.27

The data protection during data transmission from sensor nodes to the base station is a critical task and achieves minimal transmission time. The BS is responsible for generating and preserving a secret key and MIPSOE protocol can achieve a minimum packet drop ratio. In LSDAR protocol, the data is encrypted in tree level and data size increases at each level of data transmission. Hence, there is a possibility of missing a packet during data transmission. The MIPSOE protocol achieves an 8% Packet Drop Ratio with the respect to LSDAR protocol shown pictorially in Figure 5 and numerically in Table IV.

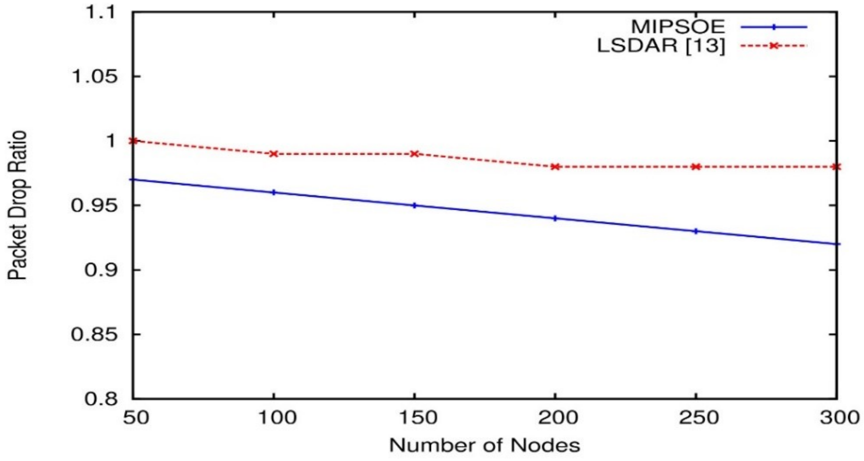

Fig. 5 Packet Drop Ratio of MIPSOE and LSDAR Protocol

Table IV Packet Delivery Ratio

Number of Nodes	Packet Delivery Ratio	
	LSDAR	MIPSOE
50	1.0	0.97
100	0.99	0.96
150	0.99	0.95
200	0.98	0.94
250	0.98	0.93
300	0.98	0.91

6 Conclusion

The MIPSOE protocol uses PSO to find the solution to formulate clusters. Based on the solution value, the fitness value is computed. Then, the cosine similarity function forms a cluster by considering similar fitness values. In the next stage, this fitness value of each node is considered to integrate with Markov Random Field (MRF) calculation. The probability of a similar eligibility node is competed to elect as Cluster Head and forms a probability set. Then, the MIPSO algorithm votes to the best node which possesses maximum value in probability set and is considered as Cluster Head. The Cluster Head aggregates data and encrypts aggregated data using XOR operation with the secret key value generated at the BS. Then, the data is forwards to the BS, the BS, in turn, decrypts data with a secret key which is stored itself and used as required in the application. Whereas, the existing protocol, LSDAR forms cluster based on varying radius size which increases energy utilization and this protocol encrypts data at each hop with the addition of extra data using XOR operation. Hence, in each step data size increases data length. Thus, LSDAR protocol pitfalls on energy utilization. Therefore, the proposed protocol simulate the result in NS2 tool and shows that 5% increased Energy Utilization, 8% improvement in End-to-End Delay, and 8% improvement in Packet Drop Ratio over LSDAR protocol.

REFERENCES

1. Prithvi S., & Sumathi S. (2020). LD^{2}FA-PSO: A Novel Learning Dynamic Deterministic Finite Automata with PSO Algorithm for Secured Energy Efficient Routing in Wireless Sensor Network. Ad Hoc Networks, 97, 1-12
2. Di Sarno C., &Garofalo A. (2014) Energy-Based Detection of Multi-layer Flooding Attacks on Wireless Sensor Network. Computer Safety, Reliability, and Security. SAFECOMP 2014, 8696, 339-349.
3. PanagiotisS., EiriniK.,&Anastasios A.,Economides. (2015). Detecting Sybil Attacks in Wireless Sensor Networks using UWB Ranging-Based Information. Expert Systems with Applications, 42(21), 7560-7572.
4. PalanN. G., BarbadekarB. V., & S. Patil. (2017). Low Energy Adaptive Clustering Hierarchy (LEACH) Protocol: A Retrospective Analysis, 2017 International Conference on Inventive Systems and Control (ICISC), Coimbatore, 1-12

5. XueyingL., Wang Y., Deng J., Zhang G., & Zhang L. (2018). Improved Particle Swarm Optimization Algorithm Based on Last-Eliminated Principle and Enhanced Information Sharing. Computational Intelligence and Neuroscience, Hindawi, 2018, 1687-5265.

6. Obado V., Djouani, Y K..& Hamam. (2012). Hidden Markov Model for Shortest Paths Testing to Detect a Wormhole Attack in a Localized Wireless Sensor Network, Procedia Computer Science, Elsevier, 10, 1010–1017.

7. XieM., HanS., TianB., &ParvinS. (2011). Anomaly Detection in Wireless Sensor Networks: A Survey.Journal of Network Computer Application, 34 1302–1325.

8. ShafieiH., KhonsariA., DerakhshiH., &MousaviP. (2014) Detection and Mitigation ofSinkhole Attacks in Wireless Sensor Networks, .Journal of Computer Sysetm, 80, 644–653.

9. Abo-Zahhad, M., Sabah M. A., Nabil S., &Shigenobu S., (2015). Mobile Sink-Based Adaptive Immune Energy-EfficientClustering Protocol for Improving the Lifetime and Stability Period of Wireless SensorNetworks. IEEE Sensors Journal, 15(8), 4576–4586.

10. M., & Kumar, P. G. (2016). An Effective Clustering Approach with Data Aggregationusing Multiple Mobile Sinks for Heterogeneous WSN. Wireless PersonalCommunications, 90(2), 423–434.

11. Thakkar, A., & Kotecha, K. (2014). Cluster Head Election for Energy and Delay Constraint Applications of Wireless Sensor Network. IEEE Sensors Journal, 14(8), 2658–2664.

12. Xiao, G., & Tan, H.-P. (2013). Clustering Algorithms For Maximizing the Lifetimefor Wireless Sensor Networks with Energy-Harvesting Sensors. Computer Networks, 57(14), 2689–2704.

13. Zhu C., Wu1S., HanG., ShuL., & WuH. (2015). A Tree-Cluster-Based Data-Gathering Algorithm for Industrial WSNs with a Mobile Sink. IEEE Access: Practical Innovations, Open Solutions, 3, 381–396.

14. Khalid H., Naveed I., Tanzila S., Amjad R., & Zahid M. (2020). LSDAR:A Light-Weight Structure Based Data Aggregation Routing Protocol with Secure Internet of Things Integrated Next-Generation Sensor Networks. Sustainable Cities and Society, 54, 1-13.