An efficient Analysis of the Fusion of Statistical-Centred Clustering and Machine Learning for WSN Energy Efficiency

Deepak Dasaratha Rao[1,*], Bala Dhandayuthapani V.[2], Ch. Subbalakshmi[3], Murlidhar Prasad Singh[4], Prashant Kumar Shukla[5], Shraddha V. Pandit[6]

[1]Indian Institute of Technology, Patna, India

[2]Department of IT, College of Computing and Information Sciences, University of Technology and Applied Sciences, Shinas campus, Oman

[3]Professor, Department of Computer Science & Engineering, Guru Nanak Instructions Technical Campus (Autonomous), Ibrahimpatnam, Ranga Reddy district, Hyderabad - 501506, Telagana State, India

[4]Assistant Professor, Department of, C.S.& E., B. P. Mandal College of Engineering, Madhepura, Bihar, India,

[5]Associate Professor (Research), Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur-522302, Andhra Pradesh, India

[6] Professor, Dept.of Artificial Intelligence and Data Science, PES Modern College of Engineering, Shivajinagar, Pune-411005, India

Emails: deepakrao@ieee.org; bala.veerasamy@utas.edu.om; subbalakshmichatti@gniindia.org; singhmurlidhar@gmail.com; prashantshukla2005@kluniversity.in; shraddha.pandit@moderncoe.edu.in

Abstract

Recently, wireless sensor networks on several challenging topics have piqued researchers' attention. Maximising a network's lifetime requires just the right combination of cluster size and number of nodes. Data transmission from nodes to cluster leaders is energy intensive, even for a modest number of clusters. If there are several clusters, many leaders will be chosen, and many nodes will rely on long-distance transmission to communicate with the home base. Therefore, in order to maximise efficiency, it is necessary to strike a balance between these two factors. WSN's major challenge is improving its energy efficiency. This is because their energy consumption defines their lifespan, and it is difficult, if not impossible, to recharge their batteries. Therefore, it is crucial to develop algorithms that consume as little energy as possible in order to maximise the network's potential. The perfect clusters are essential for the longevity of the network. Therefore, an algorithm called statistical centre energy efficient clustering approach (SEECA) is presented to increase the network's lifetime while decreasing its energy consumption. The experimental findings show that the proposed methodology SCEECA outperforms the LEACH method by a wide margin, with gains of 32% in Residual energy, 16% in Network Lifetime, and 12% in Throughput.

Keywords: Fusion Analysis, WSN; SCEECA; Throughput; LEACH.

1. Introduction:

Based on the potential placement of sensor nodes, WSN deployment can be classified as either random or deterministic. Deterministic deployment, also known as the grid deployment strategy, routinely puts sensor nodes in fixed grids. Applications that support deterministic sensor node deployment can optimise sensor node assignments to optimise target observing, connectivity in WSNs, overhead, and sensor node remaining energy [1]. Sensor node location determination is often impractical because of factors like a harsh operating environment or hostile situations. Nondeterministic (Random) networks require a complex management structure. Improved inter-network communication is a result of using machine learning to analyse sensor data. Machine learning approaches are effective and practical tools for extracting information and detecting correlations in sensor networks [2], particularly in situations in which numerical models are either not accessible or are prohibitively expensive. Using

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

ML algorithms such as supervised learning, unsupervised learning, Semi-supervised learning, Reinforcement learning, and Computational Intelligence [3] one can select the cluster head node, find the optimal path, discover useful data from collected data, discover useful data from collected data, minimize packet delay, and increase the lifetime of a WSN.

Numerous academics have found themselves drawn to the topic of energy-aware communication in wireless sensor networks since one of the primary design objectives of Wireless Sensor Networks is to maximize the network's lifetime by making the most efficient use of the energy that is available at each sensor node [4]. This has caused a large number of academics to get interested in the subject. There have been several attempts made to use models that are based on machine learning in order to keep an eye on effective energy gathering strategies for WSNs. The application of machine learning to the process of energy harvesting offers numerous advantageous outcomes [5].

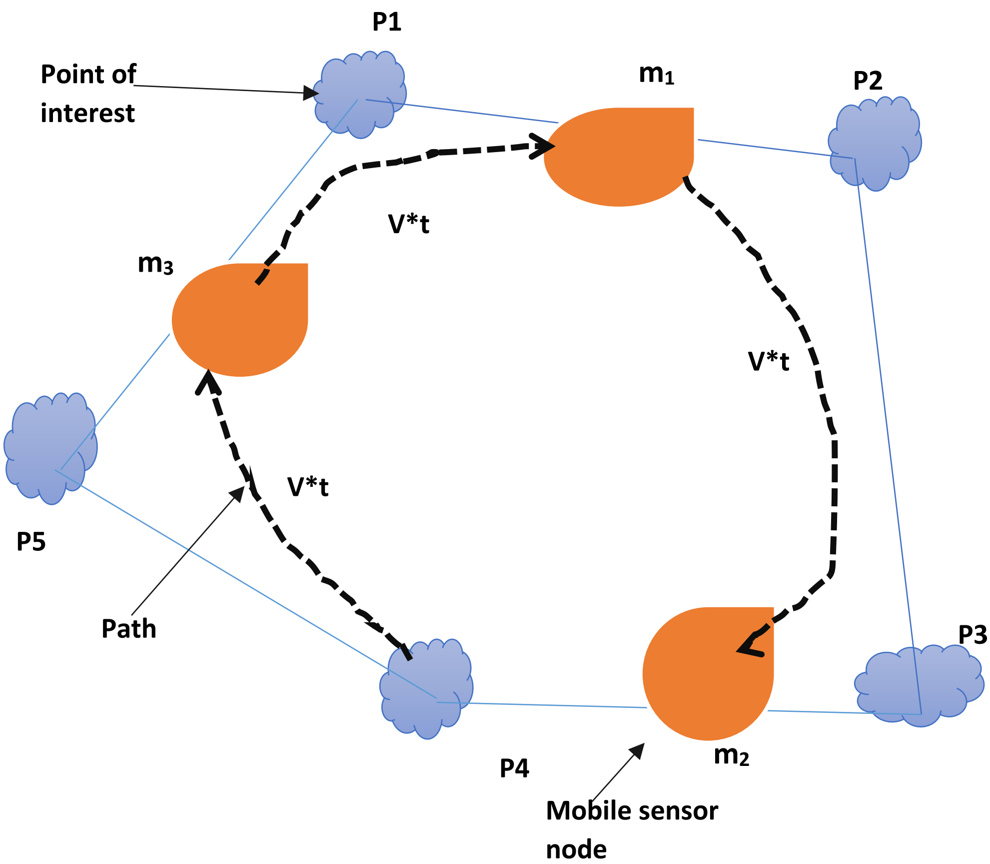

Figure 1: Mobile Sensor Nodes in Wireless Sensor Network.

A machine learning algorithm that estimates the quantity of energy that can be harvested in a given amount of time might greatly enhance the visualisation of WSNs. Energy efficiency in wireless sensor networks is difficult to achieve [6]. Because of their extreme specificity, traditional WSN approaches make it challenging for systems to adapt gradually. Responding appropriately in such situations is possible through the application of machine learning techniques. The following describe why machine learning is so crucial in WSN applications:

Sensor networks are frequently used for monitoring rapidly changing environments [7]. For example, soil degradation or sea turbulence may cause a node's position to move. Adaptable and highly functional sensor networks are desirable in these kinds of situations. WSNs can be used in exploratory contexts to collect data from remote, potentially hazardous areas. System designers may create solutions that initially do not operate as expected because of the unforeseen behavioural patterns that may come from such settings [8]. When developing accurate mathematical models to characterise the behaviour of a system, researchers often turn to WSNs for help. Some WSN tasks can be predicted with relatively elementary mathematical models, but accomplishing them often requires the employment of complex algorithms. In analogous settings, machine learning provides accurate system

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

model estimations with minimal effort [9]. Sensor network architects may have access to massive amounts of data, but struggle to make sense of it all. With limited sensor hardware resources, the WSN application often incorporates minimum data coverage criteria in addition to ensuring communication connectivity and energy sustainability. Then, machine-learning algorithms may analyse the sensor data and make recommendations about where to place sensors to get the most accurate readings [10].

WSNs have been proposed for use in a wide range of innovative applications and integrations, some of which include Cyber Physical Systems (CPS), Machine-to-Machine (M2M) communications, and Internet of Things (IoT) technologies. The overarching goal of these applications and integrations is to facilitate more intelligent decision-making and autonomous control [11]. Machine learning plays a crucial part in this situation by helping to extract the numerous levels of abstraction that are necessary to finish AI tasks with as little human intervention as possible.

2. Related Work Done:

UCS introduces a clustering method with a customizable cluster size that may be used to WSNs. The sensing field is assumed to be circular and separated into two layers. Clusters in Layer 1 are of the same size and shape, whereas those in Layer 2 come in a wide range of forms [12]. The CH should be located in the centre of the cluster to maximise efficiency. Changing the layer with a diameter close to BS shifts the concentration of a specific group, hence shifting the area covered by clusters in each layer. The developers claim that this method uses an unequal grouping strategy to achieve fair energy utilisation, which is especially helpful for networks that deal with a lot of data, and that it works fine in homogenous networks. As WSN distribution is often random, the amount of SNs per cluster might vary substantially, which is one of the approach's downsides [13].

Researchers in deterministic non-uniform node allocation show the limitations of equal clustering, which might lead to an energy hole in the network. The density of nodes increases as they go closer to the sink node in a revolutionary non-uniform deterministic node distribution model [14]. Since nodes closer to BS would be used frequently, a moderate scattered approach is recommended to balance data capacity. While the proposed method could work in predetermined node locations, the energy hole problem could arise if nodes were deployed randomly.

The authors propose EADC to distribute the network's workload uniformly despite the sensor nodes' dispersed placement. EADC produces inhomogeneous clusters to close the energy gap. The routing algorithm chooses the SN that has the most energy and the fewest hops to the sensor node in order to provide load balancing in CHs [15]. Due to the redundancy of some of the sensor nodes, additional power was required beyond what EADC could provide. Researchers solved this issue by systematically shutting down unused nodes. In addition, unnecessary sensing and transmission were disabled, which reduced overall energy use [16].

Examining nodes with diverse energy characteristics and building groups of varying sizes are two approaches to the energy hole problem. Experts agreed that the results were better than LEACH at conserving energy and extending the life of networks. By forming inhomogeneous clusters, EADUC achieves high-energy efficiency [17]. Data redundancy in overloaded regions is ignored in EADUC, leading to wasteful energy use and a shorter lifespan for the network.

The authors introduced HEED, an iterative clustering method that elects CHs based on a combination of residual energy and communication cost [18]. With HEED, the CHs are dispersed fairly over the system. After the CHs have been elected, they establish a network backbone and relay the SN's input signals to the BS via a multi-hop method. CH election in HEED consists of three key steps: initialization, major processing, and finalisation [19-21].

PEACH is a probabilistic clustering approach developed by the authors that can manage both location-aware and location-blind protocols [22-23]. PEACH employs the wireless communication property of overhearing to construct clusters, allowing for multilevel adaptive clustering without the need for additional overheads. The energy consumption model is the same as LEACH [24].

3. The Research Work Objective:

There is a need for innovative approaches to the challenges and constraints of the WSN network. Algorithms from machine learning can be used to train a network to better respond to new conditions. Energy efficiency and network lifespan are two major issues that need more study in cluster-based WSNs.

4. The Proposed Work:

The suggested algorithm calculates the intra-cluster variance, the inter-cluster variance, and the variance difference for different values of k in order to discover the optimal number of clusters. The smallest variance-

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

difference K-value is stated to be the best possible cluster size. The VDM technique, which is based on K, also successfully groups nodes into clusters. When multiple sensors in close proximity all pick up on the same events, it's helpful to group them into "clusters." The network's power is quickly depleted when every node has to communicate with the base station independently. In addition, it may lead to overloaded networks and data collisions.

The suggested SCEECA is seen in Figure 2, which also explains its essential concept. The suggested SCEECA is useful since it estimates the optimal number of clusters and performs clustering based on the values of the statistical parameters mean and variance. The Eigen values of the co-variance matrix are used to determine the cluster head.

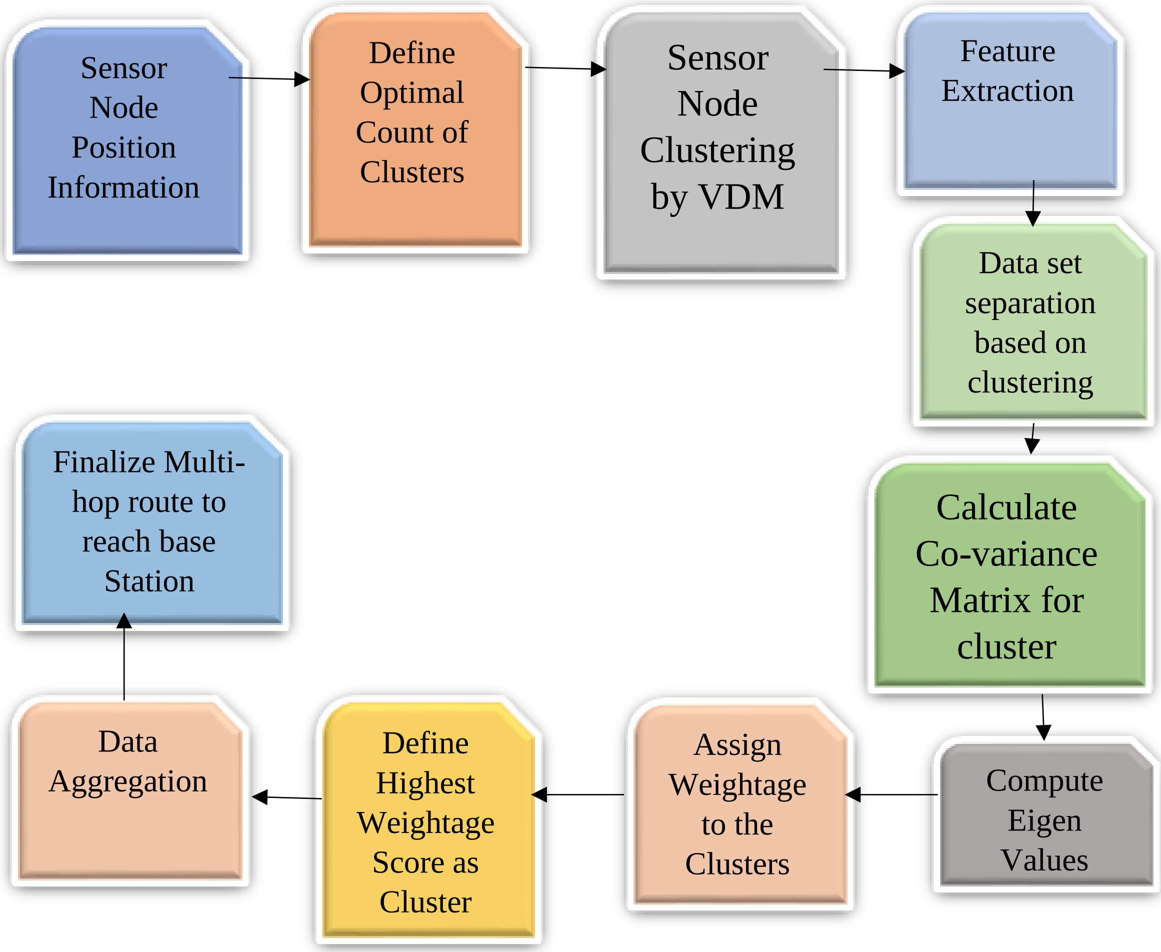

Figure 2: The Proposed Algorithm Flow chart.

The VDM technique, which is based on K, also successfully groups nodes into clusters. When multiple sensors in close proximity all pick up on the same events, it's helpful to group them into "clusters." The network's energy is rapidly depleted as individual nodes communicate with the hub. In addition, it may lead to overloaded networks and data collisions.

The learning input factor is constructed from the length of the normalized data's column vector. After that, a fixed number of iterations of backward propagation are begun. The data element is provided as a weighted input to the transfer function.

Convolution operation

$$[i]=\sum k=1Kx[i+k-1]\cdot w[k] \qquad (1)$$

where x is the input, w is the weight kernel, and y is the output.

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

$$\text{Pooling operation: } y[i]=\max(x[i:i+k-1]) \quad (2)$$

where x is the input and y is the output.

$$\text{Fully connected layer } y=\sigma(Wx+b) \quad (3)$$

where W is the weight matrix, x is the input, b is the bias vector, and σ is the activation function.

Privacy-preserving transformation: noise

$$Tpriv(xi)=xi+\epsilon\cdot noise \quad (4)$$

Homomorphic encryption operation: xaug_enci=HE(Taug(xenci))

Adversarial example generation

$$:xadv=xi+\epsilon\cdot sign(\nabla xL(Mtrained(xi,yi),yi)) \quad (5)$$

Loss function for training:

$$L(ytrue,ypred)=CrossEntropy(ytrue,ypred) \quad (6)$$

Update rule for model training:

$$Mtrained\leftarrow Mtrained-\alpha\cdot\nabla\theta L(Mtrained(xi),yi) \quad (7)$$

Data augmentation transformation:

$$Taug(x)=x+\alpha\cdot augmentation\_factor \quad (8)$$

$$\text{Privacy-preserving noise generation: } noise\sim N(0,\sigma 2) \quad (9)$$

Adversarial training objective:

$$Ladv(M,A,xi,yi)=CrossEntropy(A(M(xi)),yi) \quad (10)$$

At the subsequent level, the identical learning model discerns between malicious and benign nodes. This stage employs three distinct approaches to regulate the identification attack flow, namely "source," "destination," and "network." These approaches facilitate the differentiation of malicious nodes from regular ones, enhancing the network's ability to detect and mitigate potential security threats effectively.

Using clustering, the entire network of sensor nodes is partitioned into smaller groups, with each group electing its own CH. Each sensor communicates its discovered data by short-distance transmission to its CH, which integrates the data and communicates it via long-distance transmission to the base station. Therefore, CHs use more power to send messages than other cluster members. It is proposed that clusters vote periodically to allocate CH resources in order to achieve energy parity.

Optimal cluster size and number are essential for keeping a network running for as long as possible. It takes a lot of energy for data to be sent from cluster members to cluster leaders when there are only a few clusters. If there are many groups, many of the CH nodes will have to use long transmission to connect to the base station, and many of the cluster heads will be selected at random. In order to maximize network efficiency, it is necessary to strike a balance between these two factors. An uniform distribution of CHs is necessary for increased intra-cluster efficiency. If the cluster leaders were all chosen at the same time, the clusters wouldn't form evenly, and the network wouldn't reap the full benefits of clustering.

5. Result and Discussion:

The proposed clustering algorithm, SCEECA, and the industry standard, LEACH, were subjected to a thorough performance analysis in Mat lab. The network's characteristics and their monetary worth are listed in Table 1. The WSN configuration simulation parameters are shown in Table 2.

Table 1: Sensor Network Belongings.

Characteristic	**Assessment**
Cluster Computation	Stable
Cluster Extent	Flexible

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

Cluster Compactness	Flexible
Intra Cluster communiqué	Single-Hop
Inter Cluster communiqué	Multi-Hop
Constancy	Fixed
Flexibility	No
Node Category	Sensor
Cluster Head Assortment	Non-Probability technique
Topographies of Nodes	Position, Distance amongst node and base station, Degree, Energy Level

Table 2: Simulation Set-up details.

Constraint	**Value**
Network Size	1000m*1000m
Nodes	500
Packet Size	4000 Bits
Communication Range	20m
Initial Energy of Nodes	2 Joules
Base station Location	(1000,1000)
Transmission Energy	50 nj/bit
Communicate Amplifier Vitality	100 pj/bit/m^2
Statistics accumulation Vitality	5 nj/bit/signal

5.1. Residual Energy

Energy consumption is a major issue that needs to be considered when designing a WSN. The amount of energy lost during packet transmission is proportional to the average amount of energy required to transport a packet from a source node to its destination. Figure 3 displays the results of an LEACH analysis of the proposed system's residual energy. Because the suggested scheme selects the most stable CH in the cluster first, its typical energy consumption is less than that of LEACH, as can be shown in the diagram.

5.2. Network Lifetime

The amount of time that passes between the failure of either the first or the last node in a network is used to calculate the lifespan of the network. Additionally, this is the moment when the reserve energy of a node is depleted. The percentage of still-active nodes in each iteration is shown in Figure 4.

5.3. Throughput:

Throughput is the rate at which information may be reliably transmitted via a connection on average. Throughput is typically measured in bits per second.

Table 3: Enactment Evaluation of the Proposed Algorithm.

Number of Rounds	**Residual Energy (J)**		**Network Lifetime**		**Throughput**	
	LEACH	**SCEECA**	**LEACH**	**SCEECA**	**LEACH**	**SCEECA**
500	100	100	500	500	4	11
800	78	82	412	489	9	18
1100	65	78	349	474	12	25

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

1400	53	77	109	454	15	32
1700	29	68	115	442	15	39
2000	10	54	50	279	15	40

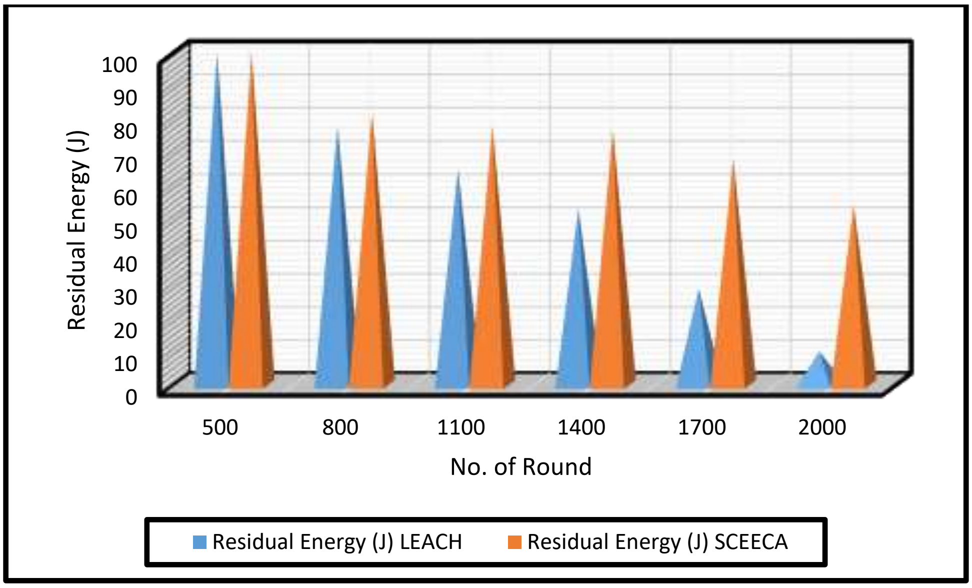

Figure 3: Residual Energy (J) Evaluation of the Proposed Algorithm.

As seen in Table 3 and Figure 3, whereas LEACH's 14% residual energy is exhausted after 2000 rounds, SCEECA's 32% residual energy is not. The proposed system's energy efficiency, relative to LEACH, thus increases linearly over time.

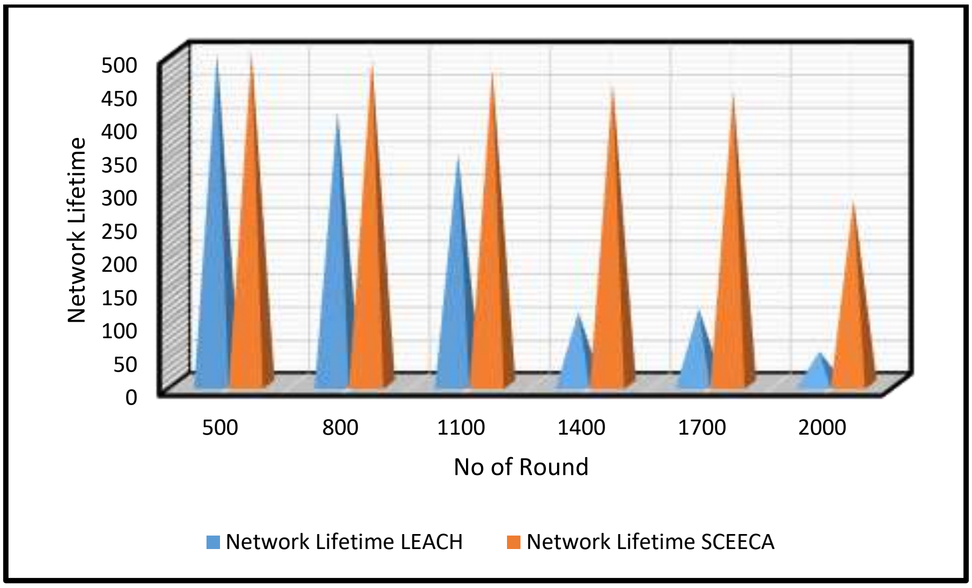

Figure 4: Network Lifetime Evaluation of the Proposed Algorithm.

According to Table 3, after 2000 iterations, LEACH has just 50 living nodes but ECCM is still going strong with 279 nodes. Figure 5 displays the results of an evaluation of the LEACH and proposed algorithms in terms of throughput and number of rounds. Table 3 demonstrates that the proposed approach outperforms the LEACH algorithm and that throughput grows linearly with the number of rounds.

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

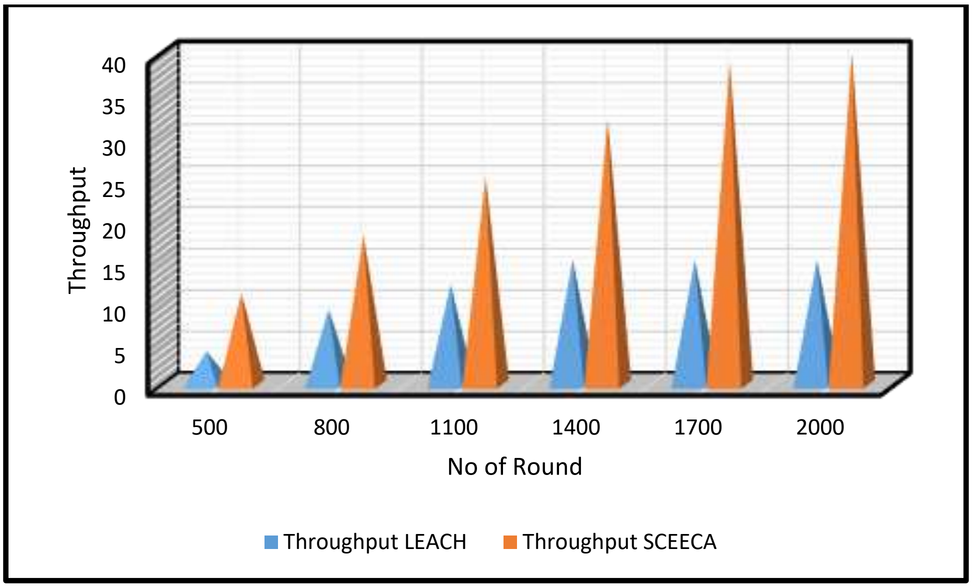

Figure 5: Throughput Evaluation of the Proposed Algorithm.

Clustering is a serious problem in constrained networks. Therefore, reducing energy use is crucial. Extensive evaluation studies reveal that SCEECA is superior to LEACH clustering in terms of reducing sensor energy consumption and maximizing system longevity. Figure 5 depicts the results of the experiments, which make it abundantly evident that the suggested approach SCEECA exceeds the LEACH technique by boosting Residual energy by 32 percentage points, Network Lifetime by 16 percentage points, and Throughput by 12 percentage points.

6. Conclusion and Future Scope:

In wireless sensor networks, sensor nodes communicate wirelessly with one another to gather information about their surroundings. The nodes in a distributed network are often low-powered and randomly placed. While the use of WSNs has increased, there are still significant limitations due to issues with storage, processing power, and battery life. Energy usage, routing algorithms, and other aspects of Wireless Sensor Networks (WSNs) are gaining traction in the academic community. Sometimes, different approaches for determining k for WSNs will produce conflicting results for the same dataset. To that end, the Variance Difference Method (VDM) was developed; it is a clustering technique that determines how many groups should be created from a given dataset. Extensive assessment experiments show that SCEECA outperforms LEACH clustering in terms of minimizing the amount of power required by sensors and maximizing the lifespan of the entire system. Figure 5 displays experimental findings that prove the suggested approach SCEECA surpasses the LEACH technique by 32 percent in terms of Residual Energy, 16 percent in terms of Network Lifetime, and 12 percent in terms of Throughput.

The supplied methods can be put to the test on any dataset derived from a practical application, and good results can be expected. Future solutions to mobility management issues may involve deep learning techniques.

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1]. Carlos-Mancilla, M, López-Mellado, E & Siller, M 2016, ‘Wireless Sensor Networks Formation: Approaches and Techniques’, Journal of Sensors, vol. 2016, pp. 1–18.

[2]. Channamma Patil & Ishwar Baidari, 2019, ‘Estimating the Optimal Number of Clusters k in a Dataset Using Data Depth’, Data Science and Engineering, vol. 4, pp. 132-140.

[3]. Chatterjee, P & Das, N 2014, ‘Coverage constrained non-uniform node deployment in wireless sensor networks for load balancing’, in Proceedings of the 1st International Conference on Applications and Innovations in Mobile Computing (AIMoC ’14), pp. 126–132.

[4]. Ding, Y, Hu, Y, Hao, K & Cheng, L 2015, ‘MPSICA: An intelligent routing recovery scheme for heterogeneous wireless sensor networks’, Information Sciences, vol. 308, pp. 49–60.

[5]. Dongyao Jia, Huaihua Zhu, Shengxiong Zou & Po Hu 2016, ‘Dynamic cluster head selection method for Wireless Sensor Networks’, IEEE Sensors Journal, vol. 16, no. 8.

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024

[6]. Jain, B, Brar, G & Malhotra, J 2018, 'EKMT-k-means clustering algorithmic solution for low energy consumption for wireless sensor networks based on minimum mean distance from base station', Networking Communication and Data Knowledge Engineering, Springer, pp. 113–123. 44.

[7]. Jan, B, Farman, H, Javed, H, Montrucchio, B, Khan, M & Ali, S 2017, 'Energy Efficient Hierarchical Clustering Approaches in Wireless Sensor Networks: A Survey', Wireless Communications and Mobile Computing, vol. 2017, pp. 1–14.

[8]. Jia, D, Zhu, H, Zou, S & Hu, P 2016, 'Dynamic cluster head selection method for wireless sensor network', IEEE Sensors Journal, vol. 16, issue 8, pp. 2746–2754.

[9]. Kingrani, Suneel Kumar, Levene, Mark & Zhang Dell, 2018, 'Estimating the number of clusters using diversity', Artificial Intelligence Research, vol. 7, issue 1, pp. 15-22.

[10]. Logambigai, R & Kannan, A 2016, 'Fuzzy logic based unequal clustering for wireless sensor networks', Wireless Networks, vol. 22, pp. 945–957.

[11]. Smys S, Abul Bashar, Wang Haoxiang (2021), "Taxonomy Classification and Comparison of Routing Protocol Based on Energy Efficient Rate", Journal of ISMAC, Vol.03, issue no.2, pp: 96-110.

[12]. Weidang Lu , Xiaohan Xu, Guoxing Huang , Bo Li , Yuan Wu , Nan Zhao and F. Richard Yu , 2021, "Energy Efficiency Optimization in SWIPT Enabled WSNs for Smart Agriculture", Ieee Transactions On Industrial Informatics, Vol. 17, issue no. 6, pp. 4335 – 4344.

[13]. Khan, A.; Khan, F. A Cost-Efficient Radiation Monitoring System for Nuclear Sites: Designing and Implementation. Intell. Autom. Soft Comput. 2022, 32, 1357–1367.

[14]. Taneja, A.; Saluja, N.; Rani, S. An energy efficient dynamic framework for resource control in massive IoT network for smart cities. Wirel. Netw. 2022, 1–12.

[15]. Deng, D.; Yuan, H. Distributed wireless sensor network system for electric field measurement. In Proceedings of the 2016 IEEE AUTOTESTCON, Anaheim, CA, USA, 12–15 September 2016.

[16]. Cui, Y.; Song, X.; Wang, C.; Wu, G.; Zhao, L. Ground-level DC electric field sensor for overhead HVDC/HVAC transmission lines in hybrid corridors. IET Gener. Transm. Distrib. 2020, 14, 4173–4178.

[17]. Rasool, F.; Drieberg, M.; Badruddin, N.; Sebastian, P.; Qian, C.T.J. Electrical battery modeling for applications in wireless sensor networks and Internet of Things. Bull. Electr. Eng. Inform. 2021, 10, 1793–1802.

[18]. He, J.; Duan, X.; Cheng, P.; Shi, L.; Cai, L. Accurate clock synchronization in wireless sensor networks with bounded noise. Automatica 2017, 81, 350–358.

[18]. Pal, A. Coverage sensitivity analysis of a wireless sensor network with different sensing range models considering boundary effects. Mater. Today Proc. 2022, 49, 3640–3645.

[19]. Jadaa, K.J.; Kamarudin, L.M.; Hussein, W.N.; Zakaria, A.; Zakaria, S.M.M.S. Multi-Target Detection and Tracking (MTDT) Algorithm Based on Probabilistic Model for Smart Cities. J. Phys. Conf. Ser. 2021, 1755, 012043.

[20]. Alobaidy, H.A.; Singh, M.J.; Behjati, M.; Nordin, R.; Abdullah, N.F. Wireless Transmissions, Propagation and Channel Modeling for IoT Technologies: Applications and Challenges. IEEE Access 2022, 10, 24095–24131.

[21]. Li, Q.; Liu, N. Monitoring area coverage optimization algorithm based on nodes perceptual mathematical model in wireless sensor networks. Comput. Commun. 2020, 155, 227–234.

[22]. Das, S.K.; Kapelko, R. On the range assignment in wireless sensor networks for minimizing the coverage-connectivity cost. ACM Trans. Sens. Netw. (TOSN) 2021, 17, 1–48.

[23]. Choi, H.H.; Lee, K. Cooperative Wireless Power Transfer for Lifetime Maximization in Wireless Multi-hop Networks. IEEE Trans. Veh. Technol. 2021, 70, 3984–3989.

[24]. Mazloomi, N.; Gholipour, M.; Zaretalab, A. Efficient configuration for multi-objective QoS optimization in wireless sensor network. Ad Hoc Netw. 2022, 125, 102730.

Doi: https://doi.org/10.54216/FPA.150217
Received: August 12, 2023 Revised: December 07, 2023 Accepted: April 17, 2024