An Energy Efficient and Scalable WSN with Enhanced Data Aggregation Accuracy

Noor Raad Saadallah[1] and Salah Abdulghani Alabady[2]

[1]*Ninevah University, Mosul, Iraq,*
[2]*University of Mosul, Mosul, Iraq*

Abstract — This paper introduces a method that combines the K-means clustering genetic algorithm (GA) and Lempel-Ziv-Welch (LZW) compression techniques to enhance the efficiency of data aggregation in wireless sensor networks (WSNs). The main goal of this research is to reduce energy consumption, improve network scalability, and enhance data aggregation accuracy. Additionally, the GA technique is employed to optimize the cluster formation process by selecting the cluster heads, while LZW compresses aggregated data to reduce transmission overhead. To further optimize network traffic, scheduling mechanisms are introduced that contribute to packets being transmitted from sensors to cluster heads. The findings of this study will contribute to advancing packet scheduling mechanisms for data aggregation in WSNs in order to reduce the number of packets from sensors to cluster heads. Simulation results confirm the system's effectiveness compared to other compression methods and non-compression scenarios relied upon in LEACH, M-LEACH, multi-hop LEACH, and sLEACH approaches.

Keywords — data aggregation, energy consumption, lossless compression, LZW algorithm, WSN, scheduling

1. Introduction

Wireless sensor networks are utilized extensively in several domains, for instance in environmental monitoring, industrial automation, and healthcare. Nevertheless, the process of collecting and analyzing data in these networks presents significant difficulties due to the large volume of data produced by the sensors. In recent years, academia has researched a range of methodologies aimed at enhancing efficiency of the data collection stage [1]. The limited memory capacity, energy, and computational capabilities of sensor nodes restrict their ability to handle massive amounts of generated data. Before transmission, the data should be compressed to help mitigate this disadvantage. Numerous data compression techniques have been proposed, mainly for the processing of images. The majority of them do not apply to sensor nodes due to issues with handling speed, energy consumption, and memory limitations.

The researchers' solution to this problem has the form of an adaptive compressed data aggregation approach that is combined with scheduling strategies, such as LZW data compression techniques. However, the task of improving information collection in WSNs is a considerable obstacle, due to the fundamental restrictions of these networks. These restrictions include power constraints, limited transmission capacity as well as the demand for power-efficient information compression processes [2].

The optimization of information-related events in WSNs holds considerable value in different applications. In ecological tracking, WSNs are used to accumulate information from varied sensing units, such as those measuring temperature, moisture or contamination levels [3]. Wearable sensing devices play a crucial role in the healthcare industry, by collecting important information for monitoring the health of patients and diagnosing specific illnesses. Optimization of manufacturing processes by deploying commercial automation solutions is greatly based on the use of sensing units. Therefore, maximizing the performance of information-related events is not solely a matter of technological value, but a reasonable necessary in all those fields [4].

This paper describes a technique that relies on K-means clustering, GA, and LZW compression to enhance the effectiveness of information collecting processes performed by sensing unit networks. This strategy intends to cluster sensing unit nodes, based on distance, using K-means clustering, enhance the collection processes by choosing collection heads with the use of GA, and compress the collected information with the help of LZW, to decrease transmission usage.

Additionally, system layouts are presented to minimize network traffic. This paper aims also to analyze the complex nature of the hybrid approach and evaluate its effectiveness in increasing the efficiency of data collection in sensor networks. The paper then evaluates the benefits and limitations of the chosen approach, with a comparative analysis of alternative methodologies already in use, including LEACH, M-LEACH, and multi-hop LEACH.

Figure 1 shows the diagram of the proposed protocol, referred to as MEECRP. K-means clustering is used to sense the nodes based on distance, thus promoting reliable information collection. This collection technique assists in arranging sensing unit nodes right into specific sets, thus maximizing information transmission and boosting network performance.

Next, GA optimization is used to pick out collection heads for enhancing the collection process, directing traffic and appropriating resources. GA assists in discovering the most appropriate solutions to complicated optimization problems.

Lastly, LZW compression is relied upon to minimize network traffic. This approach improves data transmission ef-
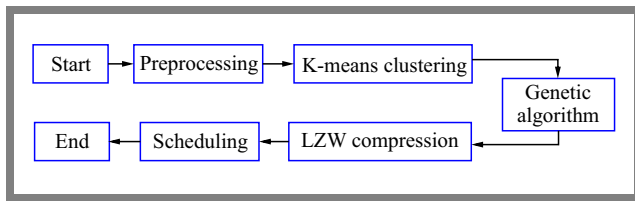
Fig. 1. Diagram of the proposed MEECRP protocol.

fectiveness, decreases energy consumption, and boosts the performance of sensing unit networks.

The performance of the proposed solution can be assessed by determining those efficiency metrics which analyze numerous elements of WSN-related procedures, such as:

- reduced power consumption, i.e. increased battery lifetime,
- network scalability, i.e. network size, coverage, and throughput,
- information collection accuracy, i.e. information aggregation efficiency, information integrity, and data compression ratio.

The paper presents a literature review in Section 2. In Section 3, the proposed model is described. In Section 4, simulation results are discussed and challenges and limitations associated with implementing the proposed method are discussed. Finally, the work is concluded in Section 5.

2. Literature Review

Recent studies reviewed in this section focus on developing the K-means cluster head selection algorithm, data packet compression algorithms, as well as on combining more than one algorithm to improve energy consumption in WSNs.

The authors of [5] proposed the LCDGRA protocol, a cluster-based WSN cooperative data-gathering and relaying method. A hybrid K-means clustering and network coding technique reduces data propagation costs and increases network performance. Simulations show that the proposed system outperforms related systems in regard to a number of various metrics.

An enhanced GA-based energy-efficient scheduling method for WSNs is presented in paper [6]. This method covers target points with a minimum number of active sensor nodes, while maintaining connectivity. The fitness function is introduced, balancing between activating a minimum number of sensor nodes, ensuring comprehensive coverage, sustaining connectivity, and selecting nodes with higher residual energy levels. The proposed method outperforms former algorithms in terms of energy efficiency, connectivity, and coverage.

Study [7] discusses trajectory data analysis using categorization and compression algorithms. The random forest classification algorithm is used, where DP, TR, and SP compression methods are tested. The role of selecting proper threshold values in achieving reliable results is stressed in that paper as well. The tradeoff between compression rate and precision level is shown. DP offers the highest compression ratio but provides lower accuracy levels, while TR has the best preci-

sion. SP is ranked in the middle. The outcomes depend also on threshold values.

Paper [8] introduced LEACH-EN, a hierarchical clustering-based routing protocol for WSNs which uses TDMA scheduling and data compression techniques to enhance energy efficiency and network scalability. This protocol reduces collisions, control overhead, and idle listening, thus achieving faster data transmission rates and extending network lifetime.

In [9], a data aggregation scheme known as F-LEACH for IoT-enabled healthcare systems is proposed. The scheme utilizes fuzzy logic to convert qualitative data to quantitative information, as well as implements complex nonlinear functions and provides approximate solutions. Simulation results show that F-LEACH outperforms similar approaches by 5–20%. The authors of [10] discussed WSN energy consumption and network longevity. The research aimed to build an expanded K-means cluster head selection (CHS) algorithm that takes onto consideration remaining energy, node density and distance to the base station. The algorithm outperforms LEACH, Mod-LEACH, and TSILEACH in terms of packet delivery ratio, throughput, network lifetime, energy usage, and node count.

Study [11] introduced EEA-CFCHS, an energy-efficient cluster creation and cluster head selection algorithm for heterogeneous WSN. The algorithm considers the energy depletion threshold for all WSN nodes. The proposed technique reduces energy consumption and increases network stability, extending its lifespan. According to simulations, the EEA-CFCHS algorithm prolongs the stability period by 42%, increases network lifespan by 62%, and extends network lifetime by 73.16% over ESRA and P-SEP protocols.

Papers [12]–[14] proved the efficacy of data aggregation in reducing energy consumption in WSNs. This is achieved by minimizing the volume of data that needs to be sent between nodes. In such a context, [15] proposed the idea of a low-electricity adaptive clustering hierarchy (LEACH). This technique aims to achieve electricity stability in the network by implementing a mechanism in which nodes periodically expect the position of cluster heads.

Article [16] expanded the application of the LEACH protocol to WSNs that encompass a mixture of sun-powered and battery-only nodes. The solar-aware low energy adaptive clustering hierarchy (sLEACH) algorithm prioritizes the selection of solar-powered nodes for the additional transmissions needed by cluster heads. [17] introduced a data aggregation strategy that integrates clustering with directed diffusion. This approach distinguishes and handles various types of data differently during the transmission process.

Paper [18] provided an elucidation of the distinctions among three clustering techniques, including LEACH and the hybrid approach. The study examined the energy-efficient distributed clustering (HEED) method and the distributed weight-based energy efficient hierarchical clustering (DWEHC) approach. The researchers also analyzed the impact of both methods on predictive data aggregation and data summarization processes. In [19], a gossip method is implemented for the purpose of
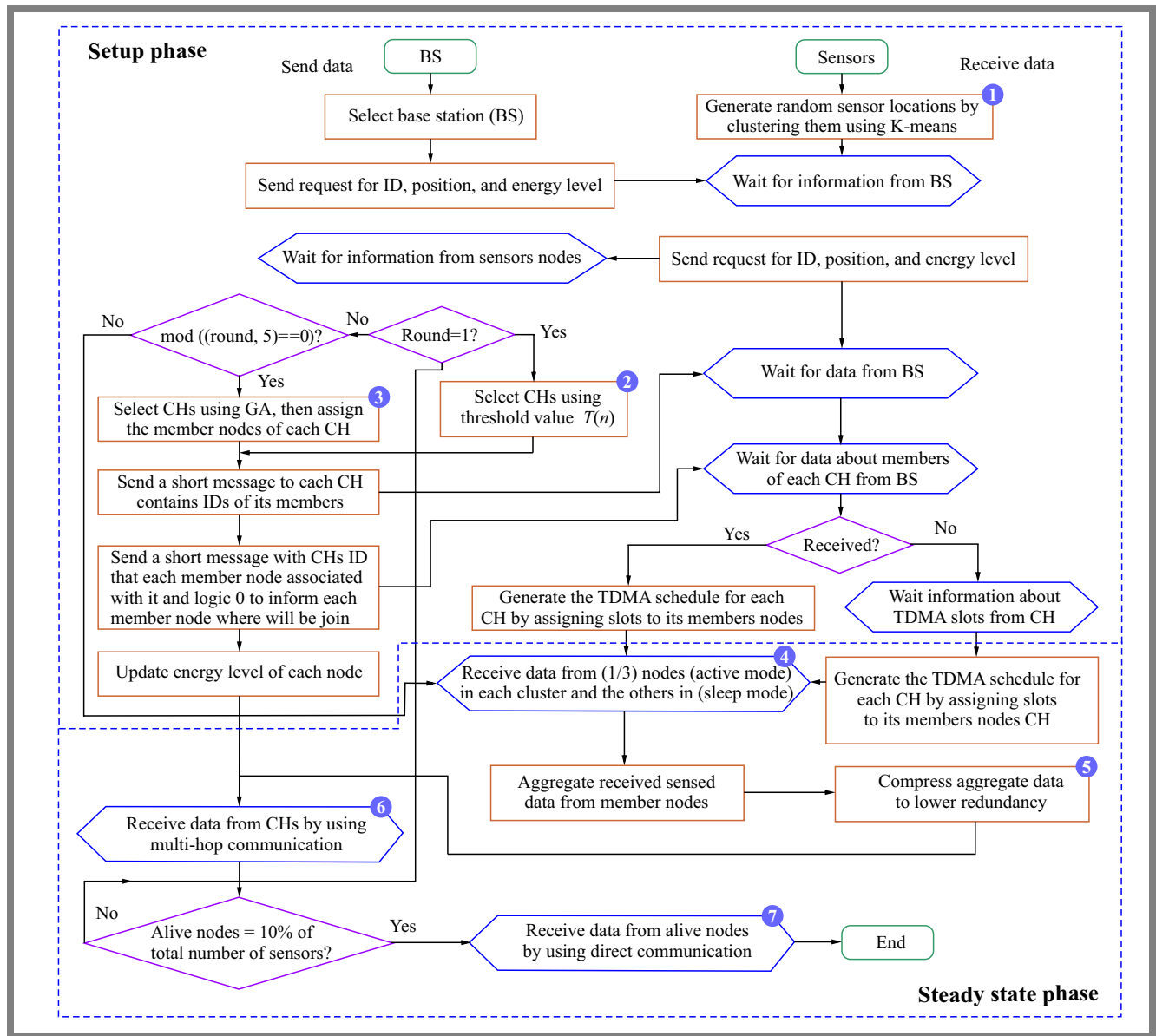
Setup phase

Fig. 2. MEECRP flow chart.

data aggregation, where data from certain nodes is not communicated directly, but is rather inferred. The authors of [20] presented a novel aggregation policy suitable for scenarios where data transmission is required in a flat topology without clusters. The authors also studied the effects of the indicator degree-based power policy (SLPC) in unscheduled wireless networks.

Investigations presented in [21] took into account channel benefits in addition to transmitting and getting energy degrees. The general FEP-gold standard power allocation approach, as proposed in [22], aims to evaluate relay-assisted variety communications by studying the body error chance (FEP).

In the evaluation of the aforementioned systems, the authors focused on improving power performance by using solar electricity for record aggregation and compression, for the duration of intervals of strength depletion or surplus. In [23], the researchers focused on energy intake-related difficulties

and community longevity of WSNs and advocated a hybrid response incorporating particle swarm optimization (PSO) and GA methodologies. To further enhance performance, the system also included a mobile sink for aggregate data from cluster heads.

Papers [24]–[26] made significant contributions in the field of energy-harvesting WSNs by developing models and approaches to estimate the availability of environmental energy and optimize its utilization by sensor nodes.

3. Proposed Model

The proposed method is divided into five steps and has been assigned the name of the multi-hop energy efficient cluster-based routing protocol (MEECRP) – Fig. 2.

3.1. Network Cluster Formation Using K-means Clustering

The K-means clustering algorithm is utilized to organize and balance sensor node clusters. The K-means clustering algorithm is a widely recognized unsupervised machine learning method that partitions data points into clusters based on their similarity. Within sensor networks, K-means clustering effectively organizes sensor nodes into cohesive clusters. The process of grouping facilitates effective data transmission management. Sharing data across sensor nodes within a given cluster may enhance efficiency by minimizing redundancy and optimizing energy use. Numerous research investigations have employed an optimization technique to reduce the distance between the sensor nodes and the base station.

The K-means algorithm offers a simpler and more efficient approach. Several methods can be employed to determine the ideal or effective number of clusters K. These methods include the gap statistic, average silhouette, and elbow techniques. The elbow technique is a valuable strategy in this context and is given by [27]:

$$SSE = \sum_{k=1}^{K} \sum_{x_i \in s_k} \| x_i - c_k \|^2 ,\qquad(1)$$

where SSE represents the sum of squared errors, x is the sensor found in clusters, and c_k is the k-th cluster. The SSE value is used to identify the ideal K number [28].

The primary objective of the K-means algorithm is to minimize the Euclidean distance, denoted as D, between the centroid of a cluster head (CH) and its cluster members (CM). Variable K is denoted by a positive integer representing the total number of clusters. The primary concept underlying K-means is the establishment of K centroids to represent each cluster. The points derived from a collection of n data sets can be effectively assigned to their nearest centroid. This represents the initial phase. Once all the relevant points have been considered, this stage concludes, and the process of grouping is finalized. Each K centroid must be recalculated as a cluster that emerges from the previous iteration. Subsequently, it is important to establish a new binding that encompasses the data set points and their nearest novel centroid. A loop may exist.

This implies that the position of the K centroid is iteratively adjusted until no further updates are required. The phases of K-means clustering are depicted in Fig. 3 [29].

3.2. Election of CH by Means of Probability Equation

The LEACH protocol designs a subset of nodes as CHs to control energy dissipation. CHs transmit group member data to the BS in one hop during the steady state phase. However, LEACH has its limitations, such as ignoring the residual energy in the process of CH selection, which leads to low-energy nodes and unbalanced clusters [30], [31]. The proposed protocol uses hierarchical clustering settings and transfers the data from the selected CHs that are elected from Eq. (2) only in the first round. These parameters are

```
Generate n random numbers between 0 and x and store
    them in an array X
X=rand (1,n)x
Generate n random numbers between 0 and y and store
    them in an array Y
Y=rand (1,n)y
Perform K-means clustering on the coordinates x and y
Set the number of clusters to k
Use default random number generator for reproducibility
Cluster the data using the K-means algorithm and obtain
    cluster assignments and centers
[idx,C] = kmeans([X',Y'], k)
Calculate the inertia (within-cluster sum of squares) of
    clustering
Inertia=sum[(X-C(idx,1)]^2+[Y-(idx,2)]^2
Calculate the silhouette score of clustering
silhouette_score=silhouette([X',Y'], idx)
Save the sensor locations and clustering results to a file
    „locations"
Save(locations, X, Y, idx, C, inertia, silhouette_score)
Visualization (optional)
Plot the clustered data with different colors and fill the
    markers
```

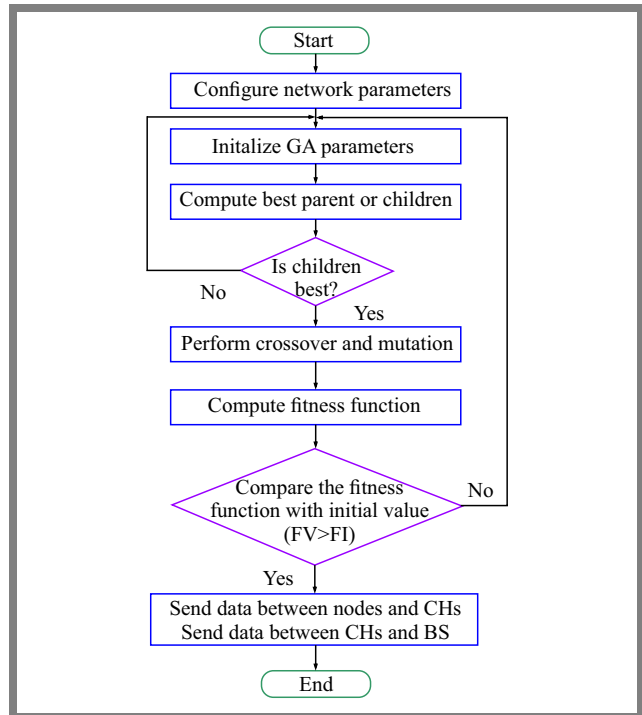**Fig. 3.** Pseudocode for clustering formation using K-means.

Fig. 4. Flow chart of the GA-based CH selection algorithm.

considered, in round one, to be initial values for GA of subsequent rounds.

$$T(n) = \begin{cases} \dfrac{p}{1 - p\left(r \bmod \frac{1}{p}\right)} & \forall\, n \in G \\ 0 & \text{if } n \notin G \end{cases}.\qquad(2)$$

3.3. Optimum Cluster Election Using GA

GAs, which draw inspiration from the mechanisms of natural selection, provide a robust approach for optimizing routing in

```
Initialize variables:
-         number of nodes n
-         number of generation generation
-         number of population numpopulation
For j =1 to generation:
    time = fitness(numpopulation, n)
    Display 'generation number (j)'
    t(j) = time
    paths(j, 1) = path_1
[ma, ind] = max(t)
best_ch = paths(ind, 1)
End for
```

Fig. 5. Pseudocode for optimum cluster election with the use of GA.

sensor networks [32]. The ability to adapt and modify routing techniques over time can result in enhanced data transfer efficiency. GAs can minimize energy usage and decrease the probability of network congestion by strategically picking the most optimal routing options. The optimal cluster head probability P_{opt} was obtained in the proposed protocol using the GA approach. Figure 4 illustrates how the cluster heads are chosen by integrating GA with our methodology.

The fitness function used in CH selection takes into account network longevity, energy consumption and load balancing when assessing the chromosomal arrangement. The evaluation includes the determination of cluster head quantities, placements, distances to base stations, cluster proportions, and distribution of load among clusters. Figure 5 shows the developed algorithm relied upon for electing optimum clusters by using GA.

3.4. Time and Event Scheduling Mechanism

Wireless sensor networks use time division multiple access (TDMA) scheduling in the LEACH protocol to improve energy efficiency and longevity. This method divides time into slots and allocates them to nodes for data transmission.

The proposed protocol assigns a group cluster head to each group to create and manage a TDMA schedule for its members. The group head combines and sends data from the group members to the base station (BS). Before starting the clustering, each node determines the cluster head. The BS receives communications about CH status, node IDs, and locations. The BS uses a GA to find the optimal probability of CH.

The BS sends an advertisement to all nodes using the optimal probability value. Each node decides its own CH role during setup and each round of the cluster. This selection uses the optimal probability of the BS and a random number. Group leaders (CHs) advise their neighbors about their situation. Non-cluster head nodes connect to the nearest cluster head node based on signal strength.

Cluster heads create and distribute a schedule for transmitting data to their constituents. Each node follows the schedule of its cluster master during each data collection cycle in the steady state phase. Nodes send data to the head of their cluster during their own time period. Group leaders collect data from group members and send it to the base station. An important achievement is that only active nodes (1/3) transmit data to the

```
If PacketType is 'Hello' then
    Set PacketSize to Model.HpacketLen
else
    Set PacketSize to Model.DpacketLen
end if
For each sender in Sender do
For each receiver in Receiver do
  Calculate distance as sqrt((Sensors(Sender).xd -
  Sensors(Receiver).xd)^2 + (Sensors(Sender).yd -
  Sensors(Receiver).yd)^2)
  Set Sensors(Sender).dis2ch to distance
  if distance > Model.do then
  Set energy loss as Model.ETX × PacketSize + Model.Emp ×
PacketSize × (distance^4)
  Subtract energy loss from Sensors(Sender).E
  if Sensors(Sender).E > 0 then
        Increment sap by 1
  end if
else
Set energy loss as Model.ETX × PacketSize + Model.Efs ×
PacketSize × (distance^2)
  Subtract energy loss from Sensors(Sender).E
  if Sensors(Sender).E > 0 then
  Increment sap by 1
      end if
  end if
end for
end for
For each receiver in Receiver do
  Set energy loss as (Model.ERX + Model.EDA) × PacketSize
  Subtract energy loss from Sensors(Receiver).E
end for
Initialize aggregatedData as an empty list
For each sender in Sender do
    if Sensors(Sender).E >0 and Sensors(Sender) has 'data' field
then
    Set compressedData as compressedData
Sensors(Sender).data
Append compressedData to aggregatedData
Set compressionEnergy as
CalculateCompressionEnergy(Sensors(Sender).data,compressed
Data,Model)
 Subtract compressionEnergy from Sensors(Sender).E
 end if
end for
For each receiver in Receiver do
 if Sensors(Receiver).E > 0 then
 Set Sensors(Receiver).data to aggregatedData
 end if
end for
For each sender in Sender do
 For each receiver in Receiver do
   if Sensors(Sender).E > and Sensors(Receiver).E>0 then
   Increment rap by 1
   end if
 end for
end for
```

Fig. 6. Pseudocode for compressed data aggregation.

CH in the collected data step. The remaining nodes are idle, but periodically switch between active, idle and sleep roles, until the number of dead nodes in the network reaches 90% of their total number. When the number of dead nodes reaches 90%, we propose another mechanism that differs from the one mentioned above. The surviving nodes do not follow the election mechanism to join the nearest group, but are sent directly to the station.

3.5. LZW Data Compression

The Lempel-Ziv-Welch compression method decreases transmission overhead by shrinking the data size. It minimizes the amount of data sent over the network, measured with the help of compression ratio, bandwidth utilization, energy consumption, transmission time, packet loss rate, and network throughput.

Data from member nodes is aggregated at the CH. The aggregation may involve averaging, summing, or choosing representative values. Then, the aggregated data is compressed using LZW and the compressed data and the data acquired from other network nodes are subsequently placed in a queue, in preparation for transmission. At regular timing intervals, the node decides to transfer data to its queue. Following each iteration, a node determines its operational state by considering an approximation of its remaining battery charge to initiate communication with other nodes in later iterations.

The active mode is selected when there is an adequate power supply needed to sustain sensing and transmitting data to the master node in the following cycle. The sleep mode is selected when the node remains active throughout the session. During this operational phase, the node is dedicated solely to the task of sensing.

The wireless module undergoes deactivation, until the subsequent round. If a node encounters power depletion within a certain round, it becomes non-functional and forfeits the sending of any data during the future round. To address this issue, nodes engage in a process of location confirmation, before initiating the routing method, thereby eliminating dead nodes.

Figure 6 illustrates the pseudocode for aggregate compressed data in the proposed protocol. Such a hybrid strategy optimizes data transmission and storage by combining these two compression methods, lowering sink data. This reduces energy usage and extends the WSN's lifetime. Hybrid compression uses LZW and aggregation compression to reduce the quantity of data and its redundancy, especially in sensor nodes that generate bulk data packets.

3.6. Communication Modes

Multi-hop WSN communication is used to send data from sensor nodes to the BS through relay nodes. This method improves network range and enhances communication-related capabilities. Multi-hop WSNs deliver compressed packets from a sensor node to the base station via a single cluster, and then CH.

In the proposed protocol, when the number of live nodes is $< 10\%$ of the total number of nodes or $<= K$ (number of clusters), the live nodes waste the remaining energy to re-elect CH, exchange control messages and send sensed data directly to the BS. These nodes, instead of being always active and sending data in each round, engage their sending mode when new events occur and when the new data differs from previous data. Such an approach is called event scheduling.
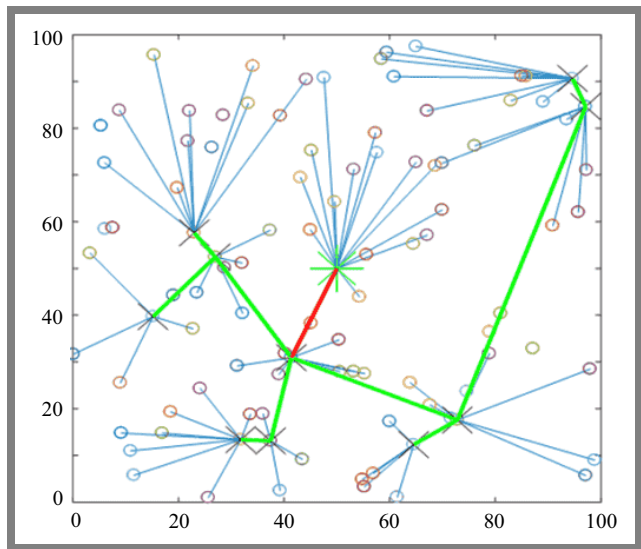
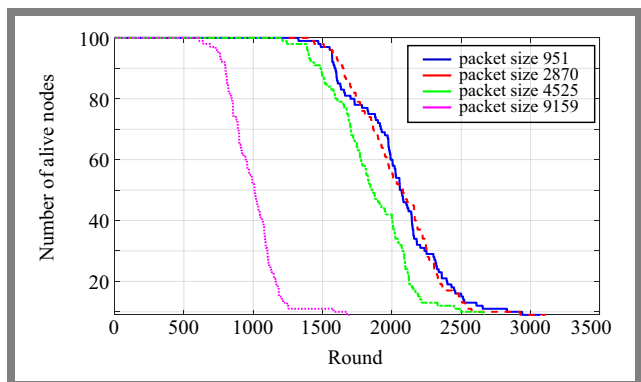**Fig. 7.** Multi-hop network used in the simulation.

Fig. 8. Network lifetime for the proposed MEECRP protocol.

4. Performance Evaluation

Protocol simulations were conducted using Matlab. Table 1 shows the parameters of the simulation model. We deployed 100 nodes randomly, with the base station located at the corner (20, 20), as shown in Fig. 7. The parameters for evaluating the offered solutions include energy consumption and the quantity of packets transmitted to the BS. The K-means algorithm is employed to create clusters after the deployment of the network, using the given parameters: an area of 100×100 m for case 1 and 300×300 m for case 2. The initial node energy level was set at 0.2 J. We simulated 4 cases with packet sizes of 951, 2780, 4525, 9159 bits, respectively, and with a total of 100 nodes. Finally, in case 3, the proposed protocol was compared with another existing routing protocol. According to the findings presented in Fig. 8, the network lifetime was achieved for $K = 25$. The energy usage per node demonstrates a positive correlation with the network's level of connectivity. This phenomenon is commonly observed as the importance of a node's connectedness becomes more prominent upon increasing the total number of nodes.

Tab. 1. Simulation parameters

| Parameter	Value
Network size	100×100, 300×300 m^2
Number of nodes	100, 200
Packet size	951, 2780, 4525, 9159 bits
Initial energy E_0	0.2 J/node
Transmitter energy E_{TX}	50 nJ/bit
Receiver energy E_{RX}	50 nJ/bit
Amplification energy for short distances ε_{fs}	10 pJ/bit/m^2
Amplification energy for long distances E_{mp}	0.0013 pJ/bit/m^2
Data aggregation energy E_{da}	0.0013 pJ/bit/m^2
Maximum number of iterations	5
Population size	10
Number of K clusters	25
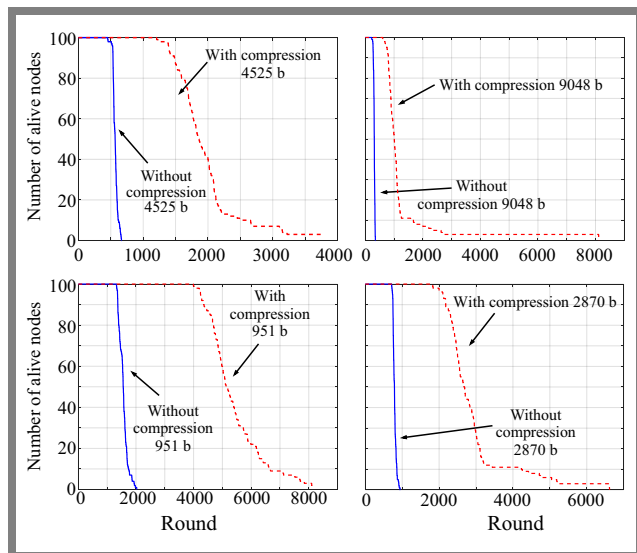
Fig. 9. Network lifetime for the proposed MEECRP approach versus LEACH protocol.

### 4.1. Simulation Results

For case 1 (100×100), Fig. 8 shows the network's lifetime after applying the proposed method to several packets of different sizes, while Fig. 9 presents the network lifetime for the proposed MEECRP approach, compared with the traditional LEACH protocol.

The impact on the lifetime of the network is positive, as the size of the data packet is changed in the LEACH protocol by applying LZW compression, but still depends on several factors, such as network configuration and node characteristics. Efficiency gains may be evaluated in terms of network lifetime and LZW data compression efficiency.
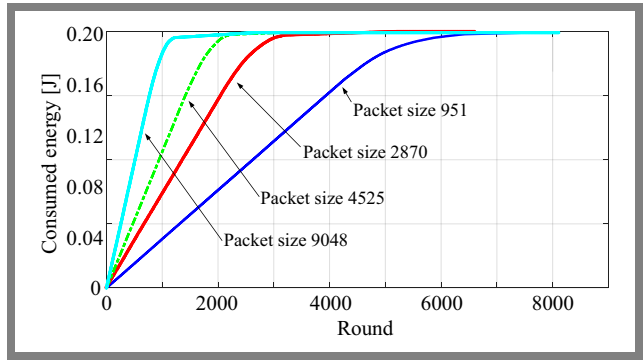
Fig. 10. Energy consumed for the proposed MEECRP approach versus packet size.
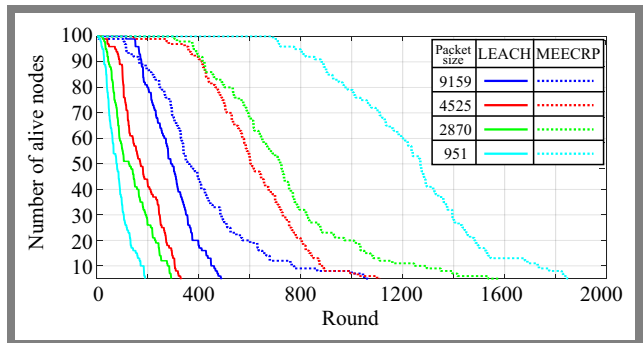
Fig. 11. Network lifetime for the proposed MEECRP approach versus the LEACH protocol for the 300×300 m case.

Figure 10 illustrates the outcomes of several simulations, where package size bits were varied. The extent of their impact on the amount of energy consumed is shown. The proposed protocol is more efficient in terms of energy consumption, as the energy reserves stored on the network's nodes play a decisive role here. While the nodes have limited power capacity, larger packets may result in faster power depletion and shorter network lifetime.

In the second case, the area of the approach field was increased to 300×300 m, while the number of nodes remained unchanged (100 nodes). The initial energy level of each node was set at 0.2 J. The base station was placed in the corner (20, 20) and the nodes were distributed randomly. The results of the simulation are shown in Fig. 11, where network lifetime for the proposed MEECRP is compared with the LEACH protocol.

Figure 12 shows the amount of energy consumed for the proposed MEECRP approach, versus the LEACH protocol. The figure shows an improvement in the efficiency of the proposed protocol, although the size of the network and node deployment density affect the distances that the data sent needs to cover. If the network is large and densely populated, larger packets may result in longer transmission distances, thus increasing power consumption.

In the third case, we compared MEECRP with other routing protocols [36]. 100 nodes were spread uniformly over an area of 100×100 m, packet size was set at 2000 bits in MEECRP. In paper [36], the packet size is 200 bits and the initial energy level for each sensor is 0.5 J.
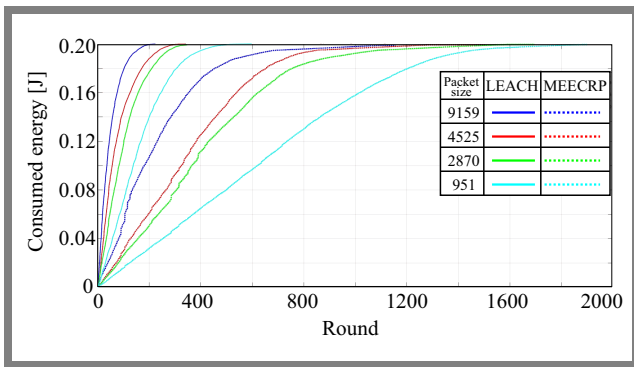
Fig. 12. Energy consumed in the proposed MEECRP approach, compared with the LEACH protocol, for various packet sizes.
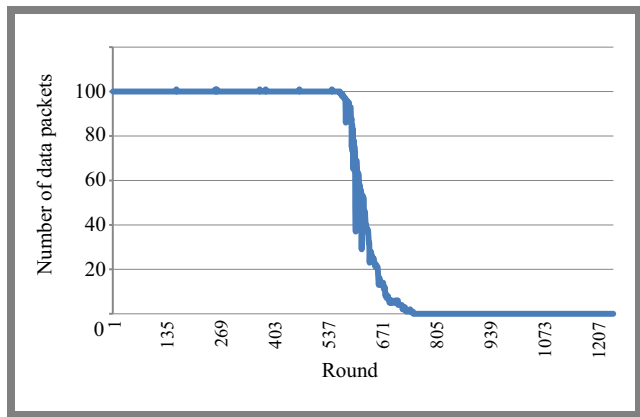

Fig. 13. Number of data packets received from BS for the MEECRP protocol (100×100 m).
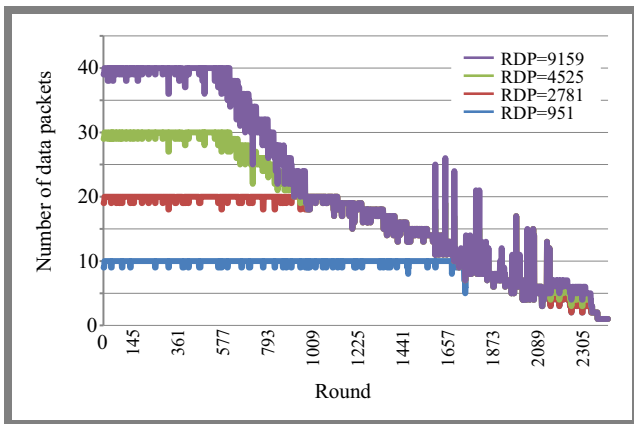

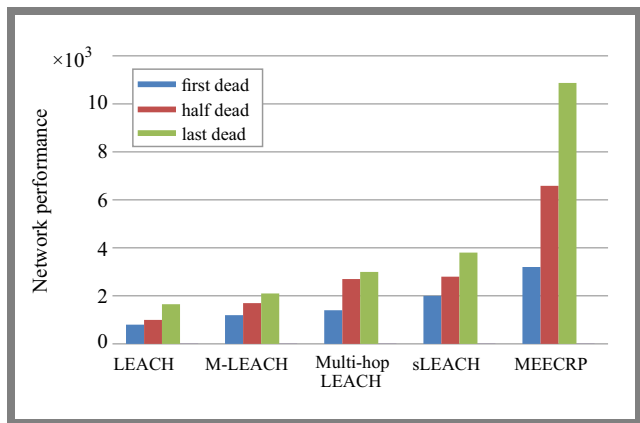**Fig. 14.** Number of data packets received from BS for the LEACH protocol (100×100 m).

Fig. 15. Network performance comparison for MEECRP and other routing protocols [36].

During the simulation, we modified the network's topology according to the real behavior of sensor nodes. Total network traffic, including data and control messages, may affect the network's lifetime. With bigger packet sizes, traffic load increases. This can be seen in Figs. 13–14, where the number of packets exchanged in the network was reduced to 40 in the proposed MEECRP protocol, but in the LEACH protocol, it equaled 100.

Even in the case of large data packets, the proposed MEECRP approach reduces the number of packets exchanged in the network, when compared with LEACH. Optimized network stability and lifetime are of utmost importance in maintaining high levels of performance. Network stability refers to the lead time from the initiation of the network to the first node failure (FND), whereas network lifetime denotes the interval between FND and the eventual failure of the last node (LND) within the network.

To evaluate the proposed solution and other approaches suggested for WSNs used in IoT, such as LEACH, M-LEACH, multi-hop LEACH, and sLEACH, a fair comparison is conducted using the above mentioned metrics.

Figure 15 shows the performance of the network relying on the proposed protocol, and compares it with solutions based on other routing protocols. The results indicate a significant improvement in network lifetime.

5. Conclusions

This research proposes a method to boost information collection capabilities of sensing networks by incorporating GA, LZW compression, and K-means approaches. The main goal of the project is to consider the impacts of those modern, cross-bred technologies on such metrics like power consumption, network scalability, and even more exact information storage. The recommended technique aims to group the sensing nodes according to distance, with the use of the K-means collection method, which certainly helps in collecting information.

In addition, GA is utilized to enhance the collection procedure by choosing collection vertices. Furthermore, LZW is applied to compress the collected information in order to lower the transmission-related burden. To further reduce network traffic, scheduling mechanisms have been introduced that lead to a reduction in the number of packets sent from sensors to cluster heads. The results of this study will contribute to the development of packet scheduling mechanisms for data aggregation in WSNs.

The results of this research will help develop methods for collecting and organizing data packets, which, in turn, will enhance the efficiency and use of resources in WSNs. As compared to various existing approaches, such as LEACH, M-LEACH, multi-hop LEACH, and sLEACH, the proposed

MEECRP approach boosts network lifetime by 100%, according to simulations. This improvement boosts the overall performance of sensing unit networks by ensuring much more efficient use of resources, enhancing information collection processes and boosting network efficiency.

Additionally, the simulations reveal that, in contrast to non-compression methods, the recommended strategy effectively decreases power usage, raises network scalability, and increases the level of precision of information gathering processes.

By using of K-means clustering, GA optimization, LZW compression, and organizing structures on several factors, which include network scalability, energy effectiveness, and statistics-amassing precision, the research sheds light on the possible benefits of incorporating those strategies in order to enhance bundle organizing in WSNs.

References

[1] S. Yoo *et al.*, "Technological Advances in Wireless Sensor Networks Enabling Diverse Internet of Things Applications", *International Journal of Distributed Sensor Networks*, vol. 14, no. 3, 2018 (https://doi.org/10.1177/1550147718763220).

[2] K. Lee and H. Lee, "An Energy-Efficient Cooperative Communication Method for Wireless Sensor Networks", *International Journal of Distributed Sensor Networks*, vol. 10, no. 3, p. 689–710, 2014 (https://doi.org/10.1155/2014/689710).

[3] L. Cao, Y. Yue, and Y. Zhang, "A Data Collection Strategy for Heterogeneous Wireless Sensor Networks Based on Energy Efficiency and Collaborative Optimization", *Computational Intelligence and Neuroscience*, vol. 2021, art. no. 9808449, 2021 (https://doi.org/10.1155/2021/9808449).

[4] M. Wu, L. Tan, and N. Xiong, "A Structure Fidelity Approach for Big Data Collection in Wireless Sensor Networks", *Sensors*, vol. 15, no. 1, pp. 248–273, 2014 (https://doi.org/10.3390/s150100248).

[5] G.P. Agbulu, G.J.R. Kumar, and A.V. Juliet, "A Lifetime-enhancing Cooperative Data Gathering and Relaying Algorithm for Cluster-based Wireless Sensor Networks", *International Journal of Distributed Sensor Networks*, vol. 16, no. 2, 2020 (https://doi.org/10.1177/1550147719900111).

[6] S. Harizan and P. Kuila, "Coverage and Connectivity Aware Energy Efficient Scheduling in Target Based Wireless Sensor Networks: An Improved Genetic Algorithm Based Approach", *Wireless Networks*, vol. 25, no. 4, pp. 1995–2011, 2019 (https://doi.org/10.1007/s11276-018-1792-2).

[7] A. Makris *et al.*, "Evaluating the Effect of Compressing Algorithms for Trajectory Similarity and Classification Problems", *GeoInformatica*, vol. 25, no. 4, pp. 679–711, 2021 (https://doi.org/10.1007/s10707-021-00434-1).

[8] D. Meenakshi and S. Kumar, "Energy Efficient Hierarchical Clustering Routing Protocol for Wireless Sensor Networks", *Advances in Computer Science and Information Technology. Networks and Communications*, pp. 409–420, 2012 (https://doi.org/10.1007/978-3-642-27299-8_43).

[9] S.N. Sajedi, M. Maadani, and M.N. Moghadam, "F-LEACH: a Fuzzy-based Data Aggregation Scheme for Healthcare IoT Systems", *Journal of Supercomputing*, vol. 78, no. 5, pp. 1030–1047, 2022 (https://doi.org/10.1007/s11227-021-03890-6).

[10] P.M. Mwangi, J.G. Ndia, and G.M. Muketha, "Cluster Head Selection Algorithms for Enhanced Energy Efficiency in Wireless Sensor Networks: A Systematic Literature Review", *International Journal of Computer Science & Engineering Survey*, vol. 13, no. 3, 2022 (https://doi.org/10.5121/ijcses.2022.13303).

[11] Q. Ren and G. Yao, "An Energy-Efficient Cluster Head Selection Scheme for Energy-Harvesting Wireless Sensor Networks", *Sensors*,

vol. 20, no. 1, art. no. 187, 2020 (https://doi.org/10.3390/s20010187).

[12] R. Rajagopalan and P.K. Varshney, "Data Aggregation Techniques in Sensor Networks: A Survey", *IEEE Communications Surveys & Tutorials*, vol. 8, no. 4, pp. 48–63, 2006 (https://doi.org/10.1109/COMST.2006.283821).

[13] D.P. Kumar, T. Amgoth, and C.S.R. Annavarapu, "Machine Learning Algorithms for Wireless Sensor Networks: A Survey", *Information Fusion*, vol. 49, pp. 1–25, 2019 (https://doi.org/10.1016/j.inffus.2018.09.013).

[14] L. Krishnamachari, D. Estrin, and S. Wicker, "The Impact of Data Aggregation in Wireless Sensor Networks", *Proc. of 22nd International Conference on Distributed Computing Systems Workshops, IEEE*, pp. 575–578, 2002 (https://doi.org/10.1109/ICDCSW.2002.1030829).

[15] M. Kaur and A. Munjal, "Data Aggregation Algorithms for Wireless Sensor Network: A Review", *Ad Hoc Networks*, vol. 100, 2020 (https://doi.org/10.1016/j.adhoc.2020.102083).

[16] M. Al-Shalabi, M. Anbar, T.-C. Wan, and A. Khasawneh, "Variants of the Low-energy Adaptive Clustering Hierarchy Protocol: Survey, Issues and Challenges", *Electronics*, vol. 7, no. 8, art. no. 136, 2018 (https://doi.org/10.3390/electronics7080136).

[17] I. Yoon, H. Kim, and D. K. Noh, "Adaptive Data Aggregation and Compression to Improve Energy Utilization in Solar-powered Wireless Sensor Networks", *Sensors*, vol. 17, no. 6, art. no. 1226, 2017 (https://doi.org/10.3390/s17061226).

[18] O. Younis and S. Fahmy, "HEED: a Hybrid, Energy-efficient, Distributed Clustering Approach for ad hoc Sensor Networks", *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004 (https://doi.org/10.1109/TMC.2004.41).

[19] P. Jesus, C. Baquero, and P.S. Almeida, "A Survey of Distributed Data Aggregation Algorithms", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 381–404, 2014 (https://doi.org/10.48550/arXiv.1110.0725).

[20] V. Freschi and E. Lattanzi, "A Study on the Impact of Packet Length on Communication in Low Power Wireless Sensor Networks under Interference", *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3820–3830, 2019 (https://doi.org/10.1109/JIOT.2019.2891841).

[21] G. Vijayaraghavan, "Intereference Management in LTE-Advanced Heteogeneous Network", Ms. Thesis, Aalto University, Finland, 2015 (https://urn.fi/URN:NBN:fi:aalto-201506303445).

[22] C. La Palombara, V. Tralli, B.M. Masini, and A. Conti, "Relay-assisted Diversity Communications", *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 415–421, 2012 (https://doi.org/10.1109/TVT.2012.2218841).

[23] N.R. Saadallah and S.A. Alabady, "Using Hybrid GA/PSO-Mobile Sink to Improve Energy Efficiency and Network Lifetime for LEACH Protocol in WSNs", *2023 IEEE 13th International Conference on System Engineering and Technology (ICSET)*, Shah Alam, Malaysia, 2023 (https://doi.org/10.1109/ICSET59111.2023.10295105).

[24] Z. Yuan *et al.*, "Energy Prediction for Energy-Harvesting Wireless Sensor: A Systematic Mapping Study", *Electronics*, vol. 12, no. 20, art. no. 4304, 2023 (https://doi.org/10.3390/electronics12204304).

[25] S. Roundy *et al.*, "Power Sources for Wireless Sensor Networks", *European Workshop on Wireless Sensor Networks*, Berlin, Germany, 2004 (https://doi.org/10.1007/978-3-540-24606-0_1).

[26] F. Mazunga and A. Nechibvute, "Ultra-low Power Techniques in Energy Harvesting Wireless Sensor Networks: Recent Advances and Issues", *Scientific African*, vol. 11, art. no. e00720, 2021 (https://doi.org/10.1016/j.sciaf.2021.e00720).

[27] B. Kumar, U.K. Tiwari, and S. Kumar, "Energy Efficient Quad Clustering Based on K-means Algorithm for Wireless Sensor Network", *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, India, 2020 (https://doi.org/10.1109/PDGC50313.2020.9315853).

[28] A. Et-Taleby, B. Mohammed, and M. Benslimane, "Faults Detection for Photovoltaic Field Based on K-means, Elbow, and Average

Silhouette Techniques through the Segmentation of a Thermal Image", *International Journal of Photoenergy*, art. no. 6617597, 2020 (https://doi.org/10.1155/2020/6617597).

[29] H. Harb *et al.*, "K-means Based Clustering Approach for Data Aggregation in Periodic Sensor Networks", *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Larnaca, Cyprus, 2014 (https://doi.org/10.1109/WiMOB.2014.6962207).

[30] T.M. Behera *et al.*, "Residual Energy-based Cluster-head Selection in WSNs for IoT Application", *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5132–5139, 2019 (https://doi.org/10.1109/JIOT.2019.2897119).

[31] H.R. Ali and H.A. Lafta, "Energy Threshold-based Cluster Head Rotation for Routing Protocol in Wireless Sensor Networks", *Journal of University of Babylon for Pure and Applied Sciences*, vol. 26, no. 7, pp. 92–106, 2018 (https://doi.org/10.29196/jubpas.v26i7.1416).

[32] N.R. Saadallah, S.A. Alabady, and F. Al-Turjman, "Energy-Efficient Cluster Head Selection via Genetic Algorithm", Al-Rafidain Engineering Journal, pp. 12–25, 2024 (https://doi.org/10.33899/rengj.2023.143955.1293).

[33] M.A. Razzaque, C. Bleakley, and S. Dobson, "Compression in Wireless Sensor Networks: A Survey and Comparative Evaluation", *ACM Transactions on Sensor Networks*, vol. 10, no. 1, pp. 1–44, 2013 (https://doi.org/10.1145/2528948).

[34] J. Jeong *et al.*, "A QoS-aware Data Aggregation in Wireless Sensor Networks", *2010 The 12th International Conference on Advanced Communication Technology (ICACT)*, 2010 (https://ieeexplore.ieee.org/document/5440486?arnumber=5440486).

[35] K.C. Barr and K. Asanović, "Energy-aware Lossless Data Compression", ACM Transactions on Computer Systems, vol. 24, no. 3, pp. 250–291, 2006 (https://doi.org/10.1145/1151690.1151692).

[36] M. Aslam *et al.*, "Survey of Extended LEACH-Based Clustering Routing Protocols for Wireless Sensor Networks", *2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems*, Liverpool, UK, 2012 (https://doi.org/10.1109/HPCC.2012.181).

Noor Raad Saadallah, Ph.D.
Computer and Information Engineering Department,
College of Electronics Engineering
https://orcid.org/0000-0002-6012-2910
E-mail: noor.saadallah@uoninevah.edu.iq
Ninevah University, Mosul, Iraq
https://uoninevah.edu.iq

Salah Abdulghani Alabady, Ph.D., Professor
Computer Engineering Department, College of Engineering
https://orcid.org/0000-0001-9687-2724
E-mail: eng.salah@uomosul.edu.iq
University of Mosul, Mosul, Iraq
https://uomosul.edu.iq