Redundancy Control Data -Driven Approach for Cluster-Based Wireless Sensor Networks

[1]Raju Imandi, [2]Pendyala Veena Vaibhavi, [3]PolimeraGunaVarshini, [4]PotnuruAnkitha, [5]Penmatsa Lavanya
[1,2,3,4,5]Dept. of CSE, Vignan's Institute of Engineering for Women, Visakhapatnam, AP, India

Abstract
Wireless Sensor Networks (WSN)is the spatially conveyed sensors utilized for checking climatic conditions. One of the major challenges in energy consumption. This paper proposes a new data aggregation technique called REDA. Redundancy Elimination Data Aggregation is a methodology used to wipe out repetitive information during all levels. The main objective to propose this algorithm is to reduce energy utilization by lesseningthe number of transmissions and increasing the efficiency of the bandwidth occupancy. REDAuses the pattern code generation approach to remove the redundancy. Thus, the transmission of redundant data within the sensor nodes as well as within the same cluster is eliminated during all iterations.

Keywords
Data Aggregation, Bandwidth Occupancy, Wireless Sensor Networks.

I. Introduction
Remote Sensor Networks (WSN), can be characterized as self-arranged and haphazardly shaped remote systems to monitor the physical conditions of the environment and to transfer their data through a network to the centralized location (also called as a base station) [1]. The sensor nodes convey among themselves through radio signals. These sensor nodes have resource constraints such as limited processing speed, storage capacity,and communication bandwidth. Since the sensor nodes should serve for a longertime, reducing energy consumption is a key issue to enhance the lifetime of the network. So, for this purpose, the usage of the energy ought to be examined at each phase of the working procedure. Thus, the researchers have examined different routing protocols, clustering schemes, data aggregation techniques [2] in which the common goal is energy advancement.

This paper is centered around building up an energy-efficient data aggregation technique. The main purpose of this technique is to reduce energy utilization usage by evacuating repetitive information. The non-repetitive is sent to the Base station through the closest group head (CH), thereby expanding the lifetime of the system. The data aggregation approach is important to reduce energy consumption and increase bandwidth utilization. Here, the information is transmitted to the local aggregator (cluster head) which joins the information from all the sensors in its group and transmits the total information to the sink.
Numerousdata aggregation protocols have been introduced for cluster-based networks. In this paper, we have used the Leach protocol for cluster head selection as well as the base idea for the Reda implementation.

II. Existing System
Leach refers to the Low Energy Adaptive Clustering Hierarchy convention, which is specifically designed as routing protocols for Wireless Sensor Network [4]. The principle point of the leach protocol is to prolong the lifetime of sensors by reducing energy consumption. Fundamentally there are three steps in this routing protocol: generation of cluster heads, the formation of clusters, communication among clusters. Leach algorithm evolved by merging the first two steps into one and the other is communication among clusters. Thus, the leach protocol has two phases.

A. Set up phase
In this phase, nodes will generate a random number between 0 and 1(including 0,1). If the random number is smaller than the threshold valueT(n), then that node will become a cluster head in this round. T(n) is calculated using the following formula:

$$T(n) = \frac{p}{1-p(r \bmod (1/p))}, N \in G$$

In the above formula
p: represents the probability of nodes to become cluster heads,
r: represents the number of rounds, N represents the number of nodes,
G: representsa set of nodes that did not become cluster heads for 1/p round.

The node that is selected as the cluster head will send the advertisement message that it has become the cluster head to its neighboring nodes.The node which receives themessage with a quality signal strengthdoesn't require much energy to transmit data with that cluster head. So, based on the signal strength the clusters are formed.
The cluster head creates ascheduled TDMA (Time Division Multiple Access), for every sensor node. According to the time gap, the nodes will send the data to the cluster head.

B. Steady Phase
In this phase, thesensor nodesstart sensing dataand send it to the cluster head for the TDMA plan. Every cluster head thencollects the data from all the sensor nodes and sends it to the base station.
The main drawback of the current frameworkis all the cluster heads send the information to the base station. For every transmission, some energy is consumed and there isa chance of transmitting redundant data. So, this paper proposes a data aggregation technique that overcomes the above problem.

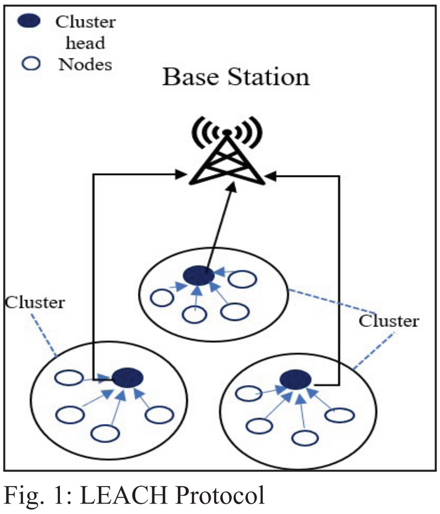


Fig. 1: LEACH Protocol

III. Proposed System

Redundancy Elimination Data Aggregation (REDA) isan approach in which the nonrepetitive data is sent to the base station. This algorithm uses the pattern code generation approach to deal with the sensed data. The sensed data is presented in the form of a set of numbers. This set of numbers is divided into different intervals. The pattern code approach is executed on all the sensor nodes. Tochoosea cluster head the LEACH protocol is used here. The selected cluster head generates the lookup table. It is then sent to all the cluster members. The lookup comprises of a set of intervals and related code.

This algorithm has primarystagesthose are t= 0 and t>0(here t represents iteration). At the initial stage, the lookup table is sent to the sensor nodes. Each sensor node compares the detected information to the values defined in the lookup table and sends it to the CH. The cluster head chooses a sensor node among them which is non -redundant. If any redundant sensor nodes are discovered, then the sensor node with a high energy level is selected. In the second iteration, as the information gathered changes the pattern code will also change. Then the sensor nodes compare their new pattern code with old pattern code. If found different they will transmit it to the CH otherwise the sensor node will drop its pattern code. In this way, the Reda eliminates the redundancy in its node as well as in neighbouring nodes.

The network is isolated into m levels. A basicsystem topology has been implemented. Generally, the sensed data can be any information such as humidity, temperature, battery voltage or moisture, etc. Right now, the parameter used is temperature.

The following diagram represents the flowchart of the proposed algorithm:

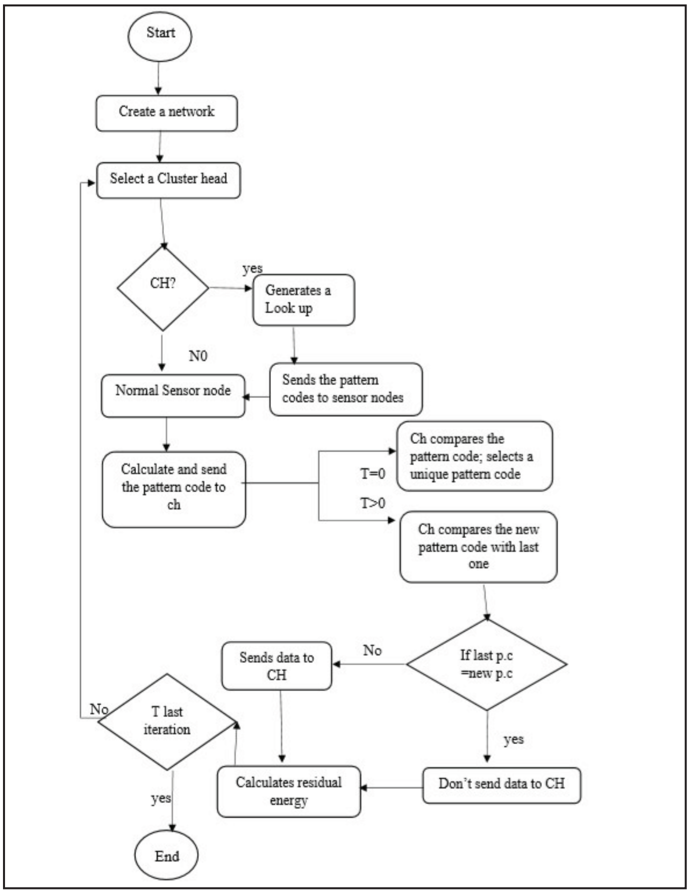
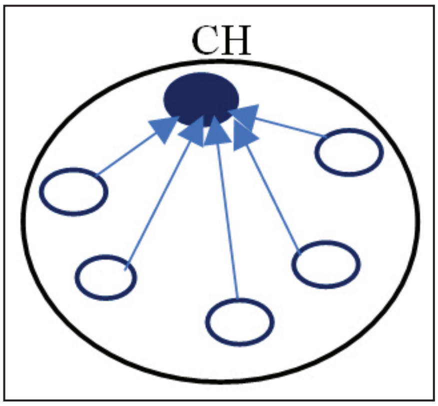


Fig. 2: REDA Flow Chart

An operating example is provided below which gives a clear idea of the functionality of the algorithm. Let us consider six sensor nodes that are deployed as a cluster. Namely SN_1, SN_2, SN_3, SN_4, SN_5, SN_6. The SN_1 is selected as the sensor node.



In this paper, the cluster head is selected using the leach protocol. The sensor node with maximum energy is chosen as the CH. Let us assume that SN1 has become the cluster head. Now, the node which is elected as the cluster head generates a lookup also known as pattern table. It consists of the intervals of temperature and the respective pattern code. Since the SN1 is cluster head it generates the following lookup table and sends it to the cluster members.

Table 1: Lookup Table for Data Intervals

Internal values	1-9	10-19	20-29	30-39	40-49
Pattern code	1	2	3	4	5

T = 0 that is at the first iteration, each sensor node has a different temperature. So, they calculate and send their pattern codes and energy levels to the cluster head. Thecluster head (SN_1) compares each sensor node's pattern code. If there are same pattern codes then it considers as redundant data. In the below table the sensor nodes {SN_6, SN_5} have the same pattern code. Similarly, the {SN_2, SN_4} has the same pattern code.The cluster head then compares the energy levels of the sensor nodes from the redundant set and chooses the one among them which has the highest energy level. If we consider the set {SN_6, SN_5}; SN_5 has the highest energy. In the same way from the set {SN_2, SN_4}; SN_2 has the highest energy. Hence at the end of the first iteration, the cluster head chooses the sensor nodes SN_2, SN_3, SN_5 to transfer the data. Then all the nodes update their energy levels based on the eq(1) and eq(2).

Table 2: Pattern Code Calculation for T = 0

Sensor id	Temperature (°C)	Pattern code	Energy levels
SN_1	6	1	0.9
SN_2	26	3	0.88
SN_3	45	5	0.89
SN_4	22	3	0.86
SN_5	34	4	0.87
SN_6	31	4	0.75

For the next iteration (T>0), again the sensor nodes gather different information and generate new pattern codes. If the new pattern

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

code is alike the old pattern code then it doesn't send it to the CH. Otherwise, it will transmit it to the CH. After that, the CH compares the received pattern codes and forms a redundant set. The node with the highest energy is selected to transfer the data.

Table 3: Pattern Code Calculation for T>0

Sensor id	Temperature (°C)	Pattern code	Energy levels
SN_1	9	1	0.88
SN_2	43	5	0.82
SN_3	44	5	0.89
SN_4	41	5	0.94
SN_5	28	3	0.87
SN_6	16	2	0.91

In Table III, the SN_3 has the same old pattern code. So, it doesn't send the pattern code to the cluster head. The other sensor nodes SN_2, SN_4, SN_5, SN_6 are eligible to disseminate their pattern codes to the CH. The CH now compares the received pattern codes and creates a redundant set. Here $\{SN_2, SN_4\}$ is the redundant set and SN_4 is chosen form the set (as it has the highest energy level). At the end of the last iteration the nodes SN_4, SN_5, SN_6 are selected to broadcast the data.

From the above example, we can clearly state that in REDA the data redundancy is eliminated by generatingpattern codes and thus minimizing the energyconsumption by reducing the number of transmissions per sensor node.

III. Simulations

System Model and Assumptions:

1. The network is divided into n levels (multi-hop communication).
2. The initial number of nodes can be n and are dispersed randomly.
3. All the nodes start with same energy level, 1 (100%)
4. The base station has no energy constraint.
5. All the nodes generate random temperatures between 1-50 °C.

In this model, the data aggregation is done by the cluster head. This collected data is sent via the cluster head that is nearest to the base station. It means the cluster heads which are close to the base station will have a high computational load. As we are using clustering here it enables bandwidth utilization efficiency and thus increases the system capacity.

The radio model that is used in this paper was introduced by the Heintzelman [4]. It considers four types of energy: energy for receiving data from the node, energy for transmitting data, energy for receiving data from another cluster and energy to transmit data to another cluster.

To calculate the energy consumed to transmit l bits message over a distancedby the transmitter and receiver is eq (1) and eq(2).

$$E_{TX} = \begin{cases} l \times E_{elect} + l \times \varepsilon_{fs} \times d^2 & \text{if} \quad d < d_0 \\ l \times E_{elect} + l \times \varepsilon_{amp} \times d^4 & \text{if} \quad d > d_0 \end{cases} \quad (1)$$

$$E_{RX} = l \times E_{elect} \quad (2)$$

Where,

d_0: refers to a Threshold distance, which is equivalent to 70 m in this model.

Eelect: refers to the electrical energy of the circuitry needed to transmit or receive l bit message.

ε_{fs} and ε_{amp}: the energy consumption factor for free space and multi-path radio models.

The energy consumption of a cluster can be seen in two different cases:

1. The CH 1 receives and aggregates data from the cluster members and sends them to the nearest cluster head CH2.
2. The CH2 relays data from the CH1, also aggregates data from its cluster members and broadcast the received data to the Base Station.

The following figure illustrates the model, the distances between CH1 and CH2 are considered to be D1 whereas the distance between CH2 to the base station is D2.

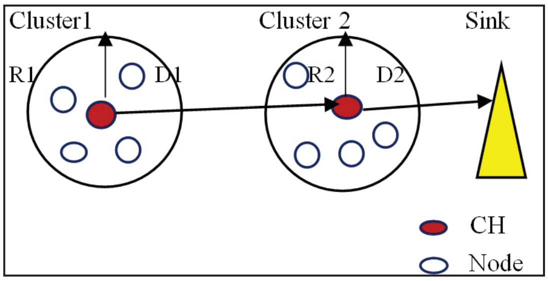


Fig. 3: Two-level Model

The total energy consumed by a CH is the sum of the energy needed for handling the internal data (its owncluster)(E_{own}) and the Energy required for relaying data from other CH's (E_{relay}). The following is the equation:

$$E_{total}^{CH} = \text{Eonn} + \text{Erelay} \quad (3)$$

In the figure 3, CH 1 has no data to rely so, the energy consumed by CH1 for transmitting 1-bit data is:

$$E_{total}^{CH\ 1} = \text{Eown} = (N_{mbr}^{1} \text{ -1}) E_{Rx}^{CH\ 1} + \gamma N_{mbr}^{1}) E_{Tx}^{CH\ 1} \quad (4)$$

Where,

$E_{Rx}^{CH\ 1}, E_{Tx}^{CH\ 1}, \gamma N_{mbr}^{1}$: Energy for transmitting and receiving 1 bit.

γ : compression factor ranging between 0 and 1

Now, considering the radius(R1) of the cluster 1 and the node density σ node:

$$E_{total}^{CH\ 1} = (\sigma\pi R_1^2 - 1)\, E_{elect} + \gamma\sigma\pi R_1^2 (E_{elect} + \varepsilon_{fs} D_1^2) \quad (5)$$

$$E_{own}^{CH2} = (N_{mbr}^{2} \text{ -1})\, E_{Rx}^{CH\ 2} + \gamma N_{mbr}^{2}) E_{Tx}^{CH\ 2}$$
$$= N_{mbr}^{2} \text{ -1})\, E_{elect} + \gamma N_{mbr}^{2})\, (Eelect + \varepsilon amp D_2^2) \quad (6)$$

$$E^{CH2}_{relay} = \gamma N^{1}_{mbr} E_{elect} + \gamma N^{1}_{mbr}(Eelect + \varepsilon amp D_2^2) \quad (7)$$

The total energy consumed by CH2 is

$$E^{CH2}_{total} = (\gamma\sigma\pi R_1^2 + \sigma\pi R_2^2 - 1)E_{elect} + \gamma\sigma\pi(R_1^2 + R_2^2)(E_{elect} + \varepsilon_{fs} D_2^2) \quad (8)$$

In this paper, the simulation has been performed on the network of 30 sensor nodes that are randomly dispersed. The Base station is located outside and doesn't possess any energy constraints. All the simulations are carried out in MATLAB.It is a programming just as in numerous specialized fields such as data analysis, problem-solvingfor experimentation and advancement of calculations.

Table 4: Configuration Parameters

Parameter	Value
Network size	400 m x 350 m
Number of nodes	30
Initial energy	32400 J
εfs	50 nJ/bit
εamp	100*pJ/bits/m*^{2}
Eelect	50.10^{-9}*J*
Transmission rate	38.4 kbps
Transmit power (Tx)	58.52 mW
Receive power (Rx)	48.6 mW
Maximum transmission range	70 m

Fig. 4 depicts the initial deployment of nodes. The nodes are deployed randomly. The x and y coordinates are used to plot the graph.

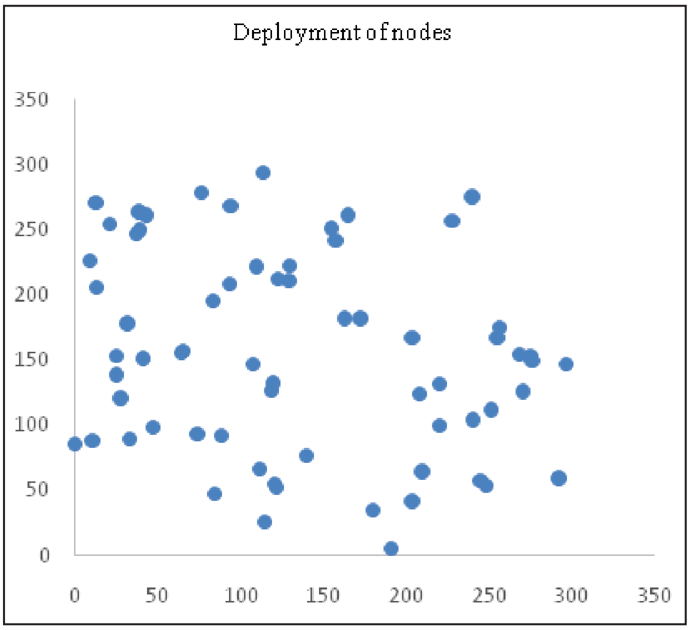


Fig. 4: Scatter Plot Showing the Deployment of Nodes

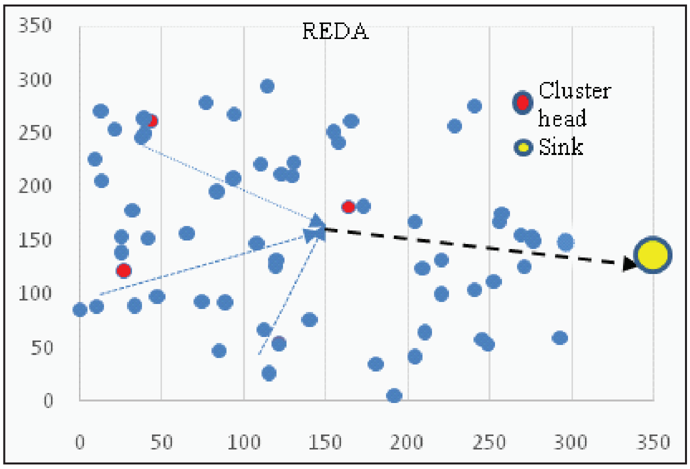


Fig. 5: The Graph Explaining the Functionality of REDA

In the above figure, the nodes which are in red are the cluster heads. The cluster heads aggregate and sends data to the nearest cluster head. The cluster head then sends it to the base station. In this way, the data is broadcasted to the base station. The blue coloured nodes are the normal nodes. The red-coloured nodes are the cluster heads whereas the Yellow colour node is the base station.

One of the advantages of REDA is all cluster heads need nearest cluster head will take the responsibility of sending the data to the Base station.

The main idea behind proposing this algorithm is energy consumption.The comparison graph of energy consumption between LEACH and REDA is presented below:

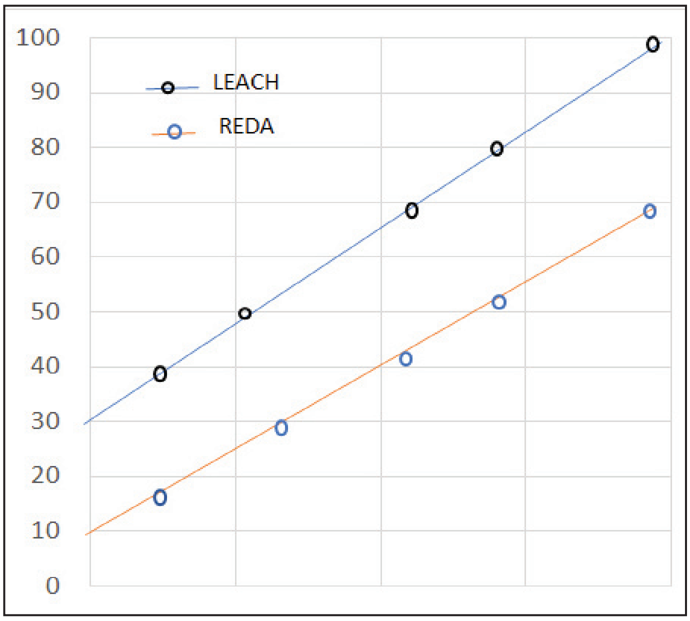


Fig. 6: energy comparison of leach and REDA graph

From the above graph, it is clear that the REDA minimizes energy consumption. The orange line indicates the REDA the blue colour indicates the LEACH. The proposed algorithm can eliminate intra-node redundancy.

IV. Conclusion

In wireless sensor networks, the data aggregation technique is one of the base approaches used to eliminate data redundancy. Even though the minimization of energy consumption is not still not

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

achieved. This paper introduces an algorithm called Redundancy Elimination Data Aggregation Technique which decreases the energy consumption in the sensor nodes as well as in the network. The main objective of this algorithm is to decrease the number of transmissions by eliminating unnecessary data.

This algorithm overcomes the problem that has been identified in the leach. Moreover, the REDA algorithm will let the sensor live for a longer time.

V. Future Scope

This algorithm further is extended to introduce an efficient algorithm for pattern code generation. In this algorithm pattern code generation approach is a tedious task. Apart from it, it should be tested for large scale networks like agriculture.

References

[1] [Online] Available: https://www.intechopen.com/books/wireless-sensor-networks-technology-andprotocols/overview-of-wireless-sensor-network
[2] [Online] Available: https://www.researchgate.net/publication/3454652_Dataaggregation_techniques_in_sensor_networks_A_survey
[3] [Online] Available: https://ssrn.com/abstract=3375325
[4] W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, "An application-specific protocol", for wireless micro sensor networks", IEEE Transactions on Wireless Communications, 1(4) pp. 660-670, 2002.
[5] [Online] Available: https://www.researchgate.net/publication/228596354_NEW_CLUSTERING_SCHEMES_FOR_WIRELESS_SENSOR_NETWORK

Veena Vaibhavi is currently pursuing her final year B.TECH in computer science and Engineering at Vignan's institute of Engineering for women. Her area of interests are wireless sensor networks and Data Science.

Varshini is currently pursuing her final year B.TECH in computer science and Engineering at Vignan's institute of Engineering for women. Her area of interests are wireless sensor networks and Artificial Intelligence.

Ankitha is currently pursuing her final year B.TECH in computer science and Engineering at Vignan's institute of Engineering for women. Her area of interests are wireless sensor networks and Cyber Security.

Lavanya is currently pursuing her final year B.TECH in computer science and Engineering at Vignan's institute of Engineering for women. Her area of interests are wireless sensor networks and cloud computing.

Raju Imandi is currently working as Assistant Professor in the department of Computer Science and Engineering at Vignan's Institute of Engineering for women. He has 4 years of teaching experience and his ares of research is wireless sensor networks and machine learning.