IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

Analysis of Energy Efficiency using Novel Algorithm Hierarchical Clustering with Map Reduce in Wireless Sensor Network Environment

S. Aravindhan[1*], D. Maruthanayagam[2]

[1]Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India
[2]Dept. of Computer Science, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India

[*]*Corresponding Author: aravindhanselvarajan@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i6.947955 | Available online at: www.ijcseonline.org

Accepted: 15/Jun/2019, Published: 30/Jun/2019

Abstract- Wireless sensor networks consist of sensor nodes, which include huge application in disaster management, habitat monitoring, military, and security preference and so on. Wireless sensor nodes might be small in size and have less-processing capability by means of low battery power consumption. These are listed as the important constraints for many WSN applications such as network lifetime, node mobility, adaptability, scalability, energy efficient, load balancing and availability. Clustering method utilized the sensor nodes is an efficient technique to achieve these goals. The different clustering algorithms also differ in their objectives. ***In this paper, a new method is to achieve the proposed technique because it supports on MAPREDUCE programming model and EM (Expected Maximization) clustering algorithm***. The key performances of the ***proposed algorithm*** **HCM (Hierarchical Clustering with MapReduce)** manage minimizing energy consumption, and take full advantage of network lifetime. The simulated performance of the results implement in the NS-2 platform, which exhibits the longer network lifetime of the ***proposed HCM algorithm and also it has performed better than the well-known clustering algorithms, DHAC*** (Distributed Hierarchical Agglomerative Clustering)***, HAC*** (Hierarchical Agglomerative Clustering)***, and K-Means with MapReduce.***

Keywords- Wireless Sensor Network, Expectation-Maximization Clustering, Cluster-Based Data Aggregation, Energy Efficient Clustering Algorithm for Maximizing Lifetime, Hierarchical Agglomerative Clustering, Distributed Hierarchical Agglomerative Clustering, K-Means Clustering using Map-Reduce Technique.

I. INTRODUCTION

A wireless sensor network (WSN) is a wireless network consisting of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions. WSN comprises the nodes in small size, low cost management, and low power consumption with limited access of memory, more computational, and communication resources and a Base Station (BS) to monitor physical or environmental conditions [1]. These nodes repeatly ***monitor environmental conditions and gathers detailed information about the physical environment*** where they are installed, after that sends the collected data to the BS. BS acts as a gateway from the sensor networks to the outside world. BS has the massive data processing capabilities and a large storage efficient. It transfers the data it receives from sensor nodes to the server from end-to-end user can access it. The sensors nodes are involved around the BS and form groups as per the need of the BS. The huge numbers of sensor nodes controls on non rechargeable batteries so as to provide the objective of load balancing, fault-tolerance, and network connectivity, grouping of nodes is needed. ***Clustering process is used for isolating sensor nodes into groups on the basis of different parameters, and chosen a group leader*** from each group [2]. The groups are known as clusters and group leaders are known as Cluster Heads (CHs) of the clusters. Parameters for forming the clusters contain distance between CH and its member, residual energy of sensor nodes, intra-cluster communication cost, and location of node with respect to BS and so forth. For the ***better optimization and maximize the network lifetime, energy consumption*** should be the major aspect of concern in WSN [3]. The hierarchical network architecture has been effectively used to afford scalable solutions in several large networks. Many hierarchical routing protocols have been ***proposed with clustering criteria, different design goals, and basic assumptions.*** Some basic challenges utilize for hierarchical routing protocol design in WSNs. Addressing those issues may possibly require new design, particularly regarding better operation and maximize the network lifetime, energy consumption**. *MapReduce is considered as a software framework for solving the clustering problems using a computing cluster***. In this paper, a new method proposed for clustering the WSN and selection of CH to each cluster. To improve the clustering technique depends upon the energy of the nodes to facilitate extend the lifetime of WSN. An approach depends upon MAP-REDUCE [1] [2] and EXPECTED MAXIMIZATION [4, 5] algorithms. ***To compare these algorithms HAC, DHAC and K-Means with MapReduce methods are evaluated and analyzed with the proposed HCM algorithm.*** The HCM algorithm is tailored for WSNs and it demonstrates experiments and simulation results and that finally concluded.

II. RELATED WORK

WSNs have been proposed various techniques for minimizing energy consumption and network life time. The aspects of hierarchical technique have been established to provide better approach to efficiently prolong the life time of WSN. ***This approach also reduces a substantial amount of energy*** which needs to be progressed upon. All the above mentioned challenges clustering have been discovered the efficient technique [6] [7]. Clustering process referred as an effective technique to improve the lifetime of WSN. **The following properties utilize with the proposed model:**

- A node is chosen as cluster head even if it has better transmission range, high residual energy, and least mobility. To achieve Energy estimation mechanism and **EA (Energy Aware) selection technique** integrated in this algorithm progress the energy performance during routing process which is built cluster by using distance and residual energy.
- For cluster production, applying the **Expectation Maximization (EM)** along with **MAP REDUCE.** This proposed method is ***partitioned into two phases*** as Mapping and Reducing. **The MAP protocol** implements mapping or assigning of sensor nodes to clusters and **REDUCE protocol** optimizes these clustering by making some changes.
- A new **CH selection mechanism implements for choosing a cluster head, which depends upon CH Rotation algorithm** for both the multi-hop and single-hop cluster models. To computing distance and residual energy of all nodes in single hop or multi hop neighbor which aids to choose CH by using CH rotation. The rotation technique is designed to assemble the constraints of dynamic environments by electing the ***CH based on main features, such as, remaining energy, energy aware distance, degree of mobility, node vulnerability consumed energy.***
- Base Station (BS) is considered as the central point of contact to the outside world and its **failure**; it can lead to total disconnection during the communication, WSN designs based on clustering. As a result, a need to discover a suitable algorithm that clusters sensor nodes in such case a BS fails and a new BS obtains the charge, new group key acquires established with minimum computation and less energy consumption. To select **EECML** more appropriate options for a clustering algorithm in order to adapt in WSN. If BS is made to fail, after that in order to continue the network in running state (as needed by critical applications) another BS will take the charge and as discussed earlier whole process of clustering is repeated and new clusters, new hierarchy and new topology will be established. A track on power consumption by sensor nodes is reserved in this complete process. ***The proposed algorithm HCM is responsible for total power consumption of all nodes before failure of BS is recorded.*** New BS can be run any random node located in any random location. The new BS is to consider different positions by repeating until the average of power consumed through all the nodes and then failure of BS in all the cases is recorded in affective way. For computing cluster stability, membership of all the nodes for clusters is recorded before and after failure of BS while clustering process is completed. These recorded readings are compared and analyzed to manage how many nodes change their membership.
- Data aggregation technique efficiently performed because ***it minimizes the number of packets to be transmitted to sink*** by aggregating the similar packets. In this paper, our attention into **cluster based data aggregation algorithms** in WSN. Data aggregation technique enhances the prolong lifetime of sensor network by decreasing the number of packets to be sent to sink or **BS.** At this point, explore the data aggregation algorithms on the basis of network topology, and then it surveys various tradeoffs in data aggregation algorithms.
- To concern about the domain problem, require to protect the energy efficient of the best path nodes to ensure their availability at all times particularly to response the urgent data packets. To **prioritize** the importance of the data packets, depending upon the urgency of the data to arrive at the destination node from the source node. And **distributed local minimization algorithm (LM)** is to compute whether switching to a different neighbor can minimize the data traffic.

2.1 Expectation-Maximization Clustering

EM (Expectation-Maximization) algorithm is used for evaluating the value of some unknown quantity, specified the values of some correlated, known quantity. ***EM is***

usually preferable to K-means because of its better convergence properties. EM is popular in statistical estimation problems involving hidden data or incomplete [8]. The EM procedure [9] is:

- The distribution parameters are initialized,
- Repeat the process until convergence:
  (a) E-Step: approximation the [E]xpected value of the unknown variables, given the current parameter approximation,
  (b) M-Step: re-approximate the distribution parameters to [M]aximize the likelihood of the data, given the estimates of the expectations of the unknown variables.

In addition the choice of initial parameters is important because the algorithm may only converge to a local optimum. In practice, poor choices can lead to bad results. Even though the convergence is guaranteed it can take a long time to achieve. Even if the missing variables and parameters do not change significantly between successive iterations then the algorithm terminates. EM algorithm applied to data clustering is very similar to the iterative K-means algorithm. **The main points of EM clustering are:**

- Initially, the cluster centers are efficiently chosen. Membership to a cluster is determined by a probability. For each point, there are as many probabilities as there are clusters. For each point the sum of probabilities equals one.
- Clusters are classified by a cluster covariance matrix and a cluster center. The covariance matrix and Cluster centers determine a Mahalanobis distance between a point and a cluster center [9].
- Covariance matrices, Cluster probabilities, and cluster centers are iteratively recomputed. For each cluster the probabilities are iteratively evaluated in the E-step.
- From the updated probabilities, all covariance matrices and cluster centers are recomputed in the M-step, so that the resulting data probability function is maximized.
- When the iteration is completed, each point is assigned to the cluster where the probability is maximal, which is equivalent to assigning the point to the ***cluster in which the Mahalanobis distance is the least.***

2.2 Energy Efficient Clustering Algorithm for Maximizing Lifetime of WSNs (EECML)

A new clustering algorithm EECML [10] is proposed by Xiang Min is designed to prolong network lifetime by minimizing the energy consumption for intra-cluster and inter-cluster communication.

CH acts as the local control center instead of frequently changing the cluster head in order to balance the load. CH is transmitting data from other CHs via multi-hop, thus the energy dissipation of the CH is much more than that of the general nodes [11, 12]. To maintain the connectivity of the entire network, it is significant that the CHs closer to the Base station maintain alive as long as possible for the inter-cluster communication. The number of the nodes in the clusters closer to the BS must be smaller than those farther away from the BS [13]. Assuming n sensor nodes are deployed in a wedge V area with angle called the clustering angle, and the nodes are deployed with uniform density (nodes/m^2). V is partitioned into m rings V_1 , V_2 , . . . , V_m . Each ring denotes a cluster, and the center distance between the two adjacent rings is d_1 , d_2 , . . . , d_m and d_i ($1<=i<=m$) is a one-hop distance for inter-cluster communication, i.e. the cluster closer to the BS is called the upper layer cluster and other is called the lower layer cluster. The schematic diagram of the model demonstrates figure 2 cluster C_1 consists of nodes located in ring V_1 , and cluster C_2 is made of nodes located in ring V_2 ,and so forth. The CH accepts data from its members and transmits fused data at once to the BS or to the upper layer cluster. The CH performs out of work before finishing its first task, when the number of the nodes in the cluster is large,. If the size of the cluster is small, then the number of nodes of the cluster is very small. The energy optimization of its nodes communicates with the support of the lower layer clusters, and its nodes may possibly have the residual energy even if the network becomes invalid, which cannot construct the best of the energy supply of its nodes, so that the optimum clustering angle should be chosen to control the size of the cluster.

In EECML, each CH acts as local control center until its working times reach the threshold. In figure 2, the size of the cluster C_m is the largest of all the clusters. Let f_i , i = 1, 2, 3, . . . , m executes **the CH (continuous working times),** which acts as the local control center. If the continuous working times of each CH with the same clustering angle satisfy $f_1 >= f_2 >= . . . . >= f_m$, after that the same clustering angle can be maintained efficiently.

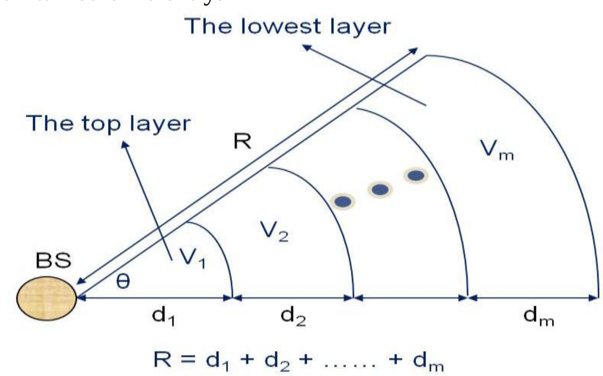

Figure 1: Algorithm model of EECML

For the energy consumption and the systems lifetime the main parameters of these algorithms are the clustering angle θ, the continuous working times f_i and one-hop distance d_{1hop}. The optimal value of these parameters is calculated in [14] by the following equations:

$$d_{opt} = \sqrt[\gamma]{\frac{2E_{elec} + E_{cpu}}{E_{amp}(\gamma - 1)}}$$
$$\theta_{opt} = \sqrt{\frac{8\pi^3(3E_{elec} + E_{cpu}}{N(2E_{elec} + E_{cpu}}}$$
$$where, d_{1hop} <= d_{opt}$$
--(1)

III. EXISTING ALGORITHMS

Some limitations for each protocol perform along with the benefits to improve and manage the functioning of protocols, which utilizes the comparison between these protocols, and additional modifications for better result [15]. Here, several clustering algorithms are used in WSN. From this section, the main consideration of ***HAC, DHAC and K-Means with MapReduce algorithms are directly related to the proposed algorithms*** based on the pre-existing hybrid algorithms.

3.1 HAC (Hierarchical Agglomerative Clustering)

HAC is a subdivision of hierarchical clustering approaches, also known as ***agglomerative technique.*** In HAC, data are not separated into certain number of groups at a single step. It consists of n clusters each cluster that controls one data object to one final cluster which includes all data objects [16]. Generally, the original raw data will be processed first and then put into a resemblance matrix for comparison. The resemblance matrix will prove the mode of processing data depending on the dissimilarity or similarity relationship between data objects. After that the data objects are categorized into clusters in accordance with the resemblance matrix, and these clusters will become larger via the multistep operation. A dendrogram can be generated the multistep operation that can utilize a visualization of the relationship between these data objects, the visualized relationship will be different when hierarchical relationships exist [17], HAC is an ***uncomplicated and effective data analysis algorithm, it has been utilized in several fields***.

3.2 DHAC (Distributed hierarchical agglomerative clustering)

DHAC is a distributed static clustering algorithm that has initiated the bottom-up scheme into WSNs clustering [18, 19]. DHAC adapts to expand using the centralized HAC technique to WSNs in a disbursed manner. As the different clustering techniques utilize top-down structure, clusters are formed before CH election in DHAC's bottom-up scheme. Using a comprehensive clustering algorithms, DHAC reveals the location information, RSS (Receive-side scaling) or connectivity data individually as the input data set for clustering process. Only local information is needed in the clustering process, in original HAC algorithm is eased. The immobility characteristic of most WSNs is exploited by DHAC to prevent re-clustering process, which results exhibits in higher energy competence. ***In the simulation, DHAC illustrates better performance than LEACH-C and LEACH in various characteristics.***

3.3 K-Means Clustering Using Map-Reduce Technique

The first step of MapReduce routines K-Means algorithm is capable of analyzing the input and output of the implementation. Input is specified as <key, value> pair in which "key" is represented as the cluster centroid and "value" is denoted as the serializable implementation of a vector in the dataset. To implement Map routine and reduce routine have two files, one must includes clusters with their centroids and other must have vectors to be clustered. Chosen centroids and the vectors to be clustered are systematized in two separate files is the initial setup for clustering data by K-means algorithm using MapReduce technique of Apache Hadoop. To achieve the algorithm to design MapReduce routines for K-means clustering. The first set of centroid is saved in the input directory of HDFS preceding to Map routine call and form the "key" field in the <key,value> pair [19,20]. The instructions needed to analyze the distance between the cluster centroid fed and the given data set as a <key,value> pair is coded in the Mapper routine. The Mapper calculates the distance between the vector value and each of the cluster centroid mentioned in the cluster set and simultaneously keeping track of the cluster to the given vector is closest. At once the computation of distances is complete the vector must be assigned to the nearest cluster. After the assignment is completed the centroid of that particular cluster is recalculated. The recalculation is completed by the ***Reduce routine and also it restructures the cluster to avoid creations of clusters*** with extreme sizes i.e. cluster having too less data vectors or a cluster having too many data vectors. As a final point, once the centroid of the given cluster is updated, ***the new set of vectors and clusters is re-written to the disk and is ready for the next iteration.***

Algorithm 1:

- procedure KmeansMapDesign
- Load Cluster file
- f_p= Mapclusterfile
- Create two list
- listnew = listold
- Call read(Mapclusterfile)
- newf_p = MapCluster()
- d_v= 0
- Assign correct centroid
- read (d_v)
- calculate centroid
- d_v = minCenter()
- Call KmeansReduce()
- end procedure = 0

Figure 2: Mapper Design for K-means Clustering

Algorithm 2:
- procedure KmeansReduceDesign
- NEW ListofClusters
- COMBINE resultant clusters from MAP CLASS
- if cluster size too high or too low then RESIZE cluster
- C_{Max} = findMaxSize(ListofClusters)
- C_{Min} = findMinSize(ListofClusters)
- if $C_{Max} > 1 / 20$
- totalSize then Resize(cluster)
- WRITE cluster FILE to output DIRECTORY
- end procedure = 0

Figure 3: Reducer Design for K-means Clustering

Algorithm 3:
- procedure Kmeans Function
- if Initial Iteration then LOAD cluster file from DIRECTORY
- else READ cluster fie from previous iteration
- Create new JOB
- SET MAPPER to map class defined
- SET REDUCER to reduce class define
- path for output DIRECTORY
- SUBMIT JOB
- end procedure = 0

Figure 4: Implementation of K-means Function

IV. THE PROPOSED HIERARCHICAL CLUSTERIN WITH MAPREDUCE (HCM) ALGORITHM

The proposed algorithm is that the base station has no constraints on its energy resources and all nodes initially have the same energy available. The operation of a sensor network begins with the cluster set up phase, in which cluster nodes will send the collected data to CH, in which clusters of the sensor nodes are formed, after that the data transmission phase. Each CH aggregates the data received from cluster nodes and relays to the BS.

a. Cluster Setup

Clustering proved to be an efficient technique to maximize the network lifetime of WSN. ***To implement a new technique is proposed for clustering the sensor nodes.*** The main motivation for proposing this system is the goodness of **EM (Expectation Maximization)** with MAP REDUCE. To improve a clustering method ***depends upon the energy*** of the nodes to maximize the lifetime of WSN. The MapReduce steps for the EM clustering are related to the K-means algorithm. EM clustering can be considering a generalized (although more complex) K-means clustering. EM clustering is also iterative like K-means. Utilizing MAP-REDUCE algorithm, the categories of key and value for proposed method are given as follows.

- key_1 → List of primary set of selected k centroids.
- $Value_1$ →List of all other nodes with their location information and energy level.
- key_2 →List of new set of k centroids.
- $Value_2$ → List of all other nodes with their cluster heads.
- key_3→List of new $k' \leq k$ centroids:
- $Value_3$→List of all other nodes form with new cluster heads.

Map Step: The maps include the Expectation step of the EM algorithm where the cluster probabilities are evaluated. In such case, the maximum probability function is computed and the minimum value is chosen that has comparable with selecting the canopy with the greatest probability. Cluster membership is determined similarly to the K-means clustering MapReduce by comparing mapped movies to the current cluster center using the Mahalanobis distance metric.

- BS→ N sensor nodes: requesting node's position and energy level.
- BS: EM (key_1, $value_1$)
- BS assigns role (Cluster Head/member) to each sensor node.
- Each cluster head transmits one hop communication about the CH and its energy level to other nodes;
- Produces output (key_1, $value_1$)

Reduce Step: The Maximization step of the EM algorithm encompass where new cluster centers are computed similarly to the K-means step.

- Read ($value_2$) /* Construct second Clustering */
- Place k' ($k' \leq k$) nodes signified as initial cluster heads
- Repeat
- If the member node is losing the energy below the threshold, it will begin searching for better CH
- Or the CH is running out of energy new CH will be assigned to the node.
- Update CH i.e. compute the mean value for each cluster
- Until no change occurred
- Produce $value_3$

Expected Maximization algorithm will be called by MAP and REDUCE.

- Base station will arbitrarily selects k nodes as initial CHs having maximum energy and closer to the node
- Repeat
- (Re) allocate each node to the cluster with the nearest CH.
- Compute the mean value of the Cluster.
- Until no change

The efficiency of clusters is computed which depends upon uniformity of node distribution.

Intra-Cluster Distance: The distance between the clusters nodes analyze its cluster centers to check whether the clusters are compact.

$$intra = \frac{1}{N}\sum_{i=1}^{K}\sum_{x \in C_i}\left|\left|x - Z_i\right|\right| \text{--------}(2)$$

Where N denotes the number of nodes in the network, K represents the number of clusters, and z_i signifies the cluster centre of cluster C_i.

Inter-Cluster Distance: This is the distance between clusters to compute this as the distance between cluster centers, and take the minimum of this value, defined as

$$inter = (||Z_i - Z_j||^2) \text{----------------}(3)$$

i=1, 2….K-1 and j=i+1….K
take only the minimum of this value.

b. Choose Cluster Heads (CHs)

CHs can be initially determined by different strategies. To prevent the issue of CH selection, many other approaches focus on how to choose appropriate CHs to attain efficient communications. From among the sensor nodes which are at the first level of distance and the next level of distance from the centroid, the nodes perform with the highest energy and choose the one which is the nearest as the CH. The completion of the clustering process by the central node transmits back the information of its cluster and individually respective CHs to each node***. As a result, each node is responsive to its cluster and CH. At this point, CHs are the nodes that satisfy below procedures:***

- The initial energy $E_{in}(n)$ of node is evaluated.
- Moreover, the distance $d(n)$ from each node to the base station or to the corresponding higher level cluster head is computed.
- In fact, evaluation of the energy needed by each node for transmission within the cluster not to higher level CH or to BS for two and three cluster formation within a cluster performs using the formula: ($E_{amp}*k*d^2$) .
- After the subsequent transmission round (maximum energy utilization) for each node is estimated and selection of CH is finished using the formula: max ($E_{in}(n)$- $E_{amp}*k*d^2$), after that the CH selection is carried out, the next cluster head selection will arise subsequent to the current round is finished.

CHs are chosen based on the Memory processing ability for each cluster and only the CH aggregate the received data from its cluster sensors and transmits to the BS.

c. CH rotation scheme

CH rotation approach performs through the entire network because its responsibilities necessitate higher energy consumption. This can be implemented rotation method, reclustering or by passing the CH role to a backup node. The choice of when to rotate the CH is handled an issue. Two frequently used strategies are periodic rotation and threshold. Periodic rotation needs precise clock synchronization between nodes, in which the threshold strategy requires to gather related information from nodes like the node residual energy, and rotate when a predefined threshold exceeded. ***A new rotation technique is proposed, that initially the centroid value nodes of each cluster operate as CH for its own cluster***. In a cluster the centroid node owning CH position up to its residual energy reaches the threshold level. If its residual energy reaches threshold level it would be send election call method packet to every node in a cluster. This CH rotation does not consume extra energy. CH rotation method uses the Chebychev Distance formula to calculate the CH with energy aware distance.

$$CH = MAX\,(|NRi_1 - Wi_1|, |NRi_2 - Wi_2|, |NRi_3 - Wi_3| \ldots . |NRi_r - Wi_r|)$$

$$Wi_n = \frac{Ri_n + Nd_n}{2}$$

Nd_n - Node distance from CH, Wi_n - weighted residual energy, NRi_n- Residual Energy of Node, Ri_n - Residual energy of node, CH_{Ri} - Cluster head residual energy.

Algorithm

Make centroid node as CH for cluster S.
If (CH energy < threshold) ∈ cluster S
{Send (ID, %energy)
Continue CH position}
Else {Do elect new CH} to cluster ∈ S
Recompute and choose CH using current cluster member-ship
Update_Cluster CH () until the stopping criterion is met

d. Data forwarding and Aggregation

Sensors transmits its own sensing data to its cluster head as each sensor knows its own cluster head and produces a shortest path to arrive its destination. **In this proposed algorithm use cluster based aggregation techniques**. Wireless sensor network is resource constraint because the sensor cannot directly send the data to the BS. In which all regular sensors can transmit data packet to a CH (local aggregator) and that ***aggregates data packet from all the regular sensors in its cluster and transmits the concise digest to the BS***. With the support of the scheme utilizes to save the energy of the sensors.

e. Priority Assignment

Data aggregation is finished by using cluster based aggregation methods. The entire data packets required to be allocated as a priority by the source node, how important is it for the packet to arrive at the destination node. ***The priority management is decided by the source node must be stored in a field, named 'Priority'***, of the data packet's header and must remain a part of the header, untampered, till it arrives the destination node. The source node utilizes the **'Priority-Assignment Table', shown in below.**

Priority level	**Importance**
Priority 1	Very High

Priority 2	High
Priority 3	Low
Priority 4	Very Low

The entire data packets consecutively possess a header field of **'Priority',** the entire path nodes know how to read the packet's priority field and take appropriate forwarding decisions. Along with the priority, the node should also recognize the current energy levels of the forwarding node, so as to decide whether the existing power status allocates it to forward the given priority data packet or not.

Priority algorithm:
START
Initialize P: Priority of data packet from packet header;
Switch (P)
Case 1: Find **ANY** suitable neighbor node for packet forwarding;
Exit switch;
Case 2: Find **ONLY** those suitable neighbor node for packet forwarding having energy level greater than 20% ;
Exit switch;
Case 3: Find **ONLY** those suitable neighbor node for packet forwarding having energy level greater than 40% ;
Exit switch;
Case 4: Find **ONLY** those suitable neighbor node for packet forwarding having energy level greater than 60% ;
Exit switch;
Case 5: Find **ONLY** those suitable neighbor node for packet forwarding having energy level greater than 80% ;
Exit switch;
Case 6: Find **ONLY** those suitable neighbor node for packet forwarding having energy level greater than 100% ;
Exit switch;
Default: Request retransmission of packet from preceding node ;
Exit Switch;
End

f. Avoid Data Traffic

The priority of data is constructed, a distributed **LM algorithm** (local minimization) is used to analyze and switching to a different neighbor can minimize the data traffic. Each node follows the given steps:

- Each node x gathers the size of the obtained data and the identity of the parents of its neighbors with ***two hops by exchanging INFO messages with its neighbors.***
- A distance less than or equal to the BS, x measures the local data traffic change if the node x changes its parent to Ni. The new local traffic is analyzed assuming node x is redirected to Ni. In such case where Ni and x encompass the same parent, the parental traffic is counted at once. If the new traffic is less than the original traffic, ***the ID of the neighbor and x records the reduced traffic size***.
- After all the neighbor measurements are complete, if the node x is not locked, it choose the neighbor and reduces the local traffic and **transmits a LOCK message** to stop it from updating its parent. If not, node x postpones its action until it is unlocked by an **UPDATE message** to continue the process. At once the LOCK message is generated by y (successfully locked y); x updates its parent to y, after that broadcasts an UPDATE message to notify its neighbors so that they can execute this algorithm with the updated information. The UPDATE message also ***unlocks y so that it can continue its own computation and update.*** If x does not obtain an acknowledgment from y after a period of time, it resumes for a random time delay and transmits a LOCK message again.

g. Energy Consumption of sensor nodes

The current BS is ended to be unsuccessful by stopping it, and another takes the charge. This arises clustering algorithm (EECML) begins forming their clusters again with similar to new location of BS. New BS can arrive at any random location, so that the three different positions execute by simulating the algorithm three times and then total power consumption of network is computed. The comparative power consumption of all sensor nodes in WSN is demonstrated in **figure 6.** The power consumption in EECML is lesser than other protocols and clarifies in graphical representation.

V. SIMULATION AND RESULTS

The proposed algorithm examined for new approach, which ***mainly minimizes consumed power of the whole network,*** which intern will **improve the network life time**. Finally the goal of work proposed algorithm ***HCM is achieved and also compare with DHAC, HAC, and K-Means with Mapreduce.*** All the simulations were executed using NS2. In this paper, the basic ***simulation settings for the simulations carried out in Table 1***. Performance metrics: (a) To **control overhead** optimization and that has the power consumption utilized for the structure formation and the route maintenance (b) **Maximize network lifetime** is defined as the duration from the beginning of the simulation to the time once one sensor runs out its energy, (c) **total number of packets received** by the mobile sink, and (d) Data delivery ratio, which is the ratio of packets successfully received by the sink to total packets transmit by source nodes. These metrics can reflect whether the hierarchical structure is efficient for data dissemination in the sensor network. (e) **Latency control** act as the time lapse between start of data dissemination from source nodes to its arrival at BS. (f) **Network scalability:** The maximum number of nodes that protocol can scale to while preserving reliable communication. (g) **Throughput:** It is the number of bytes per second received at BS.

Table 1: Parameter values

Parameters	Values
Number of Nodes	50
Simulation Area	1000×1000 (m)
Sensor Node Deployment	Random Deployment
Number of Cluster Head	5
Initial no of sink nodes	5
Transmitter Electronics (E_{TX})	50 nj/bit
Receiver Electronics (E_{RX})	
Transmit Amplifier (ε_{amp})	100 pj/bit/m^2
Traffic Type	CBR
Data Fusion Energy Dissipation Rate	5 nJ/bit
MAC Protocol	IEEE 802.11
Battery	Initial capacity is assumed to be constant
Data Rate	5 TDMA frames per 10 s
Packet size	299 bits/packet or 36 Bytes
Node Ground Speed	0.7 m/sec
Round time	35 sec

A. Energy Consumption

The energy consumption analyzed in the form of total energy consumed over time for these four protocols in Figure 5. Graph of energy has illustrated energy consumption Y-axis in scale of 10 joules and on X-axis time per 100 seconds. It is evident from the graph plotted for three popular available protocols for clustering and the HCM one. The blue line depicts that the proposed algorithms is better than the three standards available in the network optimization. Quantitatively the saved energy may perform like meager but for sensors this energy is considered to be very high energy as it can maximize the network life time by hours depending upon the application and the environment of the network.

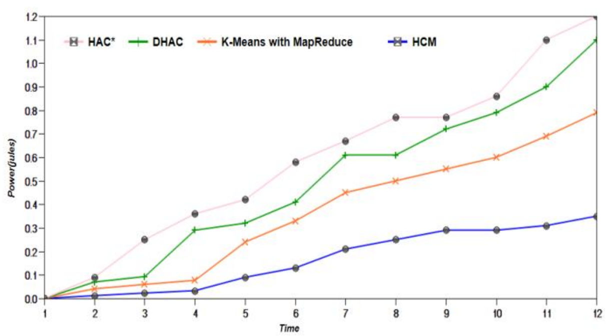

Figure 5: comparison of consumption

B. Network Lifetime

Network lifetime reveals due to energy aware selection technique and the selection of pathway with maximal nodal surplus energy in figure 6, the ***HCM algorithm performed better than other than algorithm in terms of network life time and packet delivery ratio.*** Using this algorithm will not pressure the sensor nodes with less residual energy so that the critical nodes avoid from depleting their energy earlier and prevent route rediscovery for each route break. The HCM clustering mechanism is to be simulated and the enhancement of the HCM routing algorithm together with the clustering algorithm is to be analyzed. To analyze the enhanced performance in terms of less significant energy consumption expect because the lifespan of the fusion node is maximized the performance of the network lifetime.

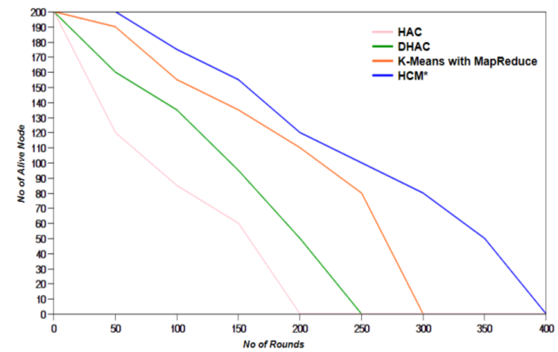

Figure 6: Network lifetime graph

C. Energy Efficiency

In terms of energy efficiency of the three algorithms defined as the number of sensor nodes transmission times per unit energy in Figure 7. ***HCM algorithm provides the best energy efficiency, while DHAC, HAC, and K-Means with MapReduce give the worst efficiency.*** The result illustrates that HCM algorithm is the most efficient scheme and the transmission data per unit energy is delivered ***up to approximate 50% in the end.***

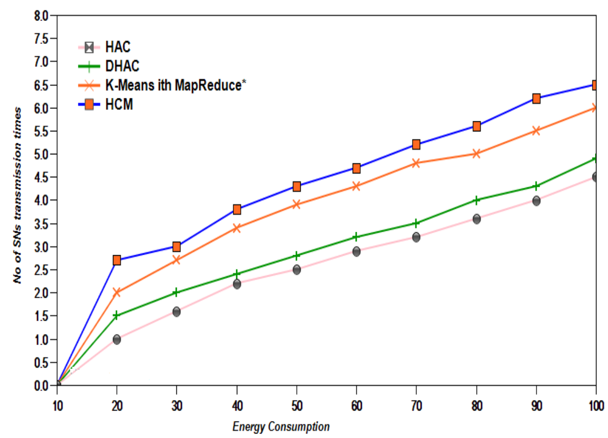

Figure 7: Energy efficiency

VI. CONCLUSION

WSNs have inspired their versatile (adaptable) applications attention of researchers. It has been addressing problems related to WSN. **Clustering process plays significant responsibilities of a WSN.** Several clustering protocols have been proposed and involved to solve problems like load balancing, scalability issue, and spatial reuse. From this paper, proposed an innovative and energy efficient clustering hybrid multipath routing algorithm with an efficient clustering technique. ***The implementation of the proposed algorithm HCM handles with other standard***

algorithms. The experimental results show that ***HCM algorithm outperforms HAC, DHAC and K-Means with MapReduce in all criteria***. All the simulation results reveal that **HCM algorithm provides higher energy efficiency** to congregate the constraints of WSNs. In addition, HCM algorithm is **more flexible in terms of different WSN application scenarios.**

REFERENCES

[1]. Foto N Afrati and Jeffrey D Ullman, "Optimizing joins in a map-reduce environment", In Proceedings of the 13th International Conference on Extending Database Technology. ACM, 99–110, 2010.

[2]. Prajesh P Anchalia, Anjan K Koundinya, and NK Srinath, "MapReduce design of K-means clustering algorithm", In Information Science and Applications (ICISA), 2013 International Conference on. IEEE, 1–5, 2013.

[3]. Xiaoli Cui, Pingfei Zhu, Xin Yang, Keqiu Li, and Changqing Ji, "Optimized big data K-means clustering using MapReduce", The Journal of Supercomputing 70, 3 (2014), 1249–1259, 2014.

[4]. A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", Blackwell Publishing, Oxford, England, UK, 1977.

[5]. I. D. Dinov, "Expectation Maximization and Mixture Modeling Tutorial", University of California, Los Angeles, USA, 2008

[6]. Liliana M Arboleda C, Nidal Nasser, "Cluster- based Routing Protocol for Mobile Sensor Networks", In: 3rd International Conference on Quality of Service in Heterogeneous Wired/ Wireless Networks, Waterloo, Canada, 2006.

[7]. Ratish Agarwal, Mahesh Motwani, "Survey of Clustering Algorithm for MANET", In: International Journal on Computer Science and Engineering, Vol.12, pp.98-104, 2009.

[8]. Celeux, G., Govaert, G, "A classification EM algorithm for clustering and two stochastic versions", Comput. Stat. Data Anal. 14, 315–332 (1992)

[9]. Dempster, A., Laird, N., Rubin, "D.: Maximum likelihood for incomplete data via the EM algorithm", J. Roy. Stat. Soc. 39(B), 1–38 (1977)

[10]. Sangho Yi, Junyoung Heo, Jiman Hong, "PEACH: Power-efficient and Adaptive Clustering Hierarchy Protocol for Wireless Sensor Networks", Science Direct, Computer communications 30, pages 2842-2852, 2007.

[11]. Guo B, Li Z, "United voting dynamic cluster routing algorithm based on residual-energy in wireless sensor networks", Journal of Electronics & Information Technology 29(12), pages 3006-3010,2007.

[12]. Yuan H-y, Yang S-q, Li X-l et al, "Time-controlled routing algorithm for sensor networks", Journal of System Simulation 20(5), pages 11631166, 2008.

[13]. Li C-f, Cheng G-h, Ye M. et al, "An uneven cluster-based routing protocol for wireless sensor networks", Chinese Journal of Computers 30(8), pages 2730, 2007.

[14]. M. Aslam, N. Javaid, A. Rahim, U. Nazir, A. Bibi, Z. A. Khan,"Survey of extended LEACH based clustering routing protocols for wireless sensor network", IEEE transaction on Antennas and propagation, Vol. 50, N. 5 May 2012.

[15]. G.Karypis, E.H.Han and V.Kumar, "CHAMELEON: Hierarchical clustering using dynamic modeling", IEEE Computer, 32, pp. 68-75, 1997.

[16]. J.A.S. Almeida, L.M.S. Barbosa, A.A.C.C. Pais and S.J. Formosinho, "Improving Hierarchical Cluster Analysis: A new method with outlier detection and automatic clustering", Chemo metrics and Intelligent Laboratory Systems, 87, pp. 208-217, 2007.

[17]. C. H. Lung and C. Zhou, "Using hierarchical agglomerative clustering in wireless sensor networks: An energy-efficient and flexible approach," Ad Hoc Networks, vol. 8, pp. 328-344, 2010.

[18]. C. Zhou, "Application and Evaluation of Hierarchical Agglomerative Clustering in Wireless Sensor Networks," MASc Thesis, Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, 2008.

[19]. C. Romesburg, "Cluster analysis for researchers: Lulu", com, 2004.

[20]. Y. Zhao and G. Karypis, "Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering," Machine Learning, vol. 55, no. 3, pp. 311-331, June 2004.

Authors Profile

Dr.D.Maruthanayagam received his **Ph.D** Degree from Manonmaniam Sundaranar University, Tirunelveli in the year 2014. He received his **M.Phil** Degree from Bharathidasan University, Trichy in the year 2005. He received his **M.C.A** Degree from Madras University, Chennai in the year 2000. He is working as **Head & Professor**, PG and Research Department of Computer Science, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India. He has **18 years** of experience in academic field. He has published **5 books**, **33 papers** in International Journals and **30 papers** in National & International Conferences so far. His areas of interest include Computer Networks, Grid Computing, Cloud Computing and Mobile Computing.

S.Aravindhan received his **M.Phil** degree from Thiruvalluvar University, Vellore in the year 2012. He has received his **MCA** degree from Anna University, Chennai in the year 2011. He is **pursuing** his **Ph.D** degree at Sri Vijay Vidyalaya College of Arts & Science (Affiliated Periyar University), Dharmapuri, Tamilnadu, India. His areas of interest include Data Mining, Cloud Computing and Computer Networks.

