Cluster Based Data Aggregation Protocol to improve the network life time of Wireless Sensor Networks

[1]S.Sujanthi,[2] Dr.Nithya Kalyani
[1]Assistant Professor,[2] Associate Professor
[1]Department of Computer Science and Engineering,[2] Department of Information Technology
[1]Ariyalur Engineering College,[2] K.S.R College of Engineering
[1]Ariyalur, [2] Tiruchengode, India

***Abstract*:** In recent years, energy efficiency and data gathering is a major concern in many applications of Wireless Sensor Networks (WSNs). One of the important issues in WSNs is how to save the energy consumption for prolonging the network lifetime. For this purpose, many novel innovative techniques are required to improve the energy efficiency and lifetime of the network. The paper proposes the sink mobility and nodes heterogeneity aware cluster-based data aggregation algorithm (MHCDA) for efficient bandwidth utilization and an increase in network life time. The proposes algorithm uses a predefined region for the aggregation of packets at the cluster head for minimizing computation and communication cost. and Bandwidth Efficient Heterogeneity Aware Cluster based Data Aggregation (BHCDA) algorithm presents the solution for the effective data gathering with in-network aggregation. It embodies the optimal approach by Intra and inter-cluster aggregation on the randomly distributed nodes with variable data generation rate while routing data to sink.

IndexTerms ***—Data aggregation; MHCDA; BHCDA; Energy consumption; Wireless Sensor Network.***

I. INTRODUCTION

Internet of Things (IoT) in correlation with WSN supports for the distributed event based an application that differs from conventional IP application. Emerging technology as wire-less communications and IoT has paid attention for device to device communications. The low cost sensing devices has further scope in the evolution and applications of WSNs–IoT [1].

Data aggregation algorithms in WSNs–IoT are application specific and export the data to data base as shown in Fig.1. Wireless sensor networks (WSNs) with IoT provide a dynamic network infrastructure for data aggregation, distribution, and processing. It aims to combine several packets from the sensor nodes in an efficient manner and compress the data [2]. The CH acts as a gateway for receiving information from the nodes and applies the additive and divisible aggregation functions to reduce the packet count. This saves the energy, improves the network lifetime, and bandwidth utilization. In the hierarchical WSN, resource allocation primarily relates to the amount of bandwidth given to the CH.

This paper considers the network model that refers to cluster-based aggregation with node heterogeneity in-terms of energy and sinks mobility. With static environmental conditions, the layered architecture used in cluster-based data aggregation increases the network lifetime and reduces the energy consumption. But with the mobile environment,

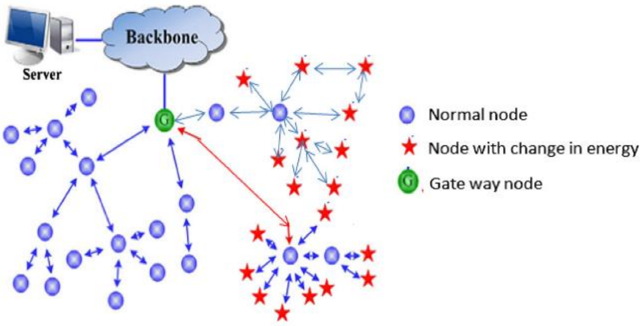

Fig1. Heterogeneous network of WSN

It is challenging to aggregate the data. The proposed MHCDA uses multi-hop aggregations to reduce the energy consumption and increase the network lifetime. It exploits the aggregation functions by the correlation of number of packets and data generated by the nodes at variable rate. It effectively reduces the packet count reaching to the mobile sink, and hence, improves bandwidth utilization with reduced packet delivery ratio (PDR) and Throughput, while transferring the data to other data bases [3].

This paper proposed a novel BHCDA to reduce the number of transmissions of data packets from source nodes to the mobile sink. Mainly it focuses on the symmetric aggregation functions at cluster head, which are perfectly compressible. The aggregation function considers the correlation of the random data generated by cluster member nodes in the cluster region. Each node in the network generates data by use of random function with the standard deviation for data in the range of 0 and 1. The performance of BHCDA is measured by aggregating the variable rate of data packets generated with consideration of throughput, energy consumption, and packet delivery ratio with bandwidth utilization. BHCDA shows significant improvement in throughput and packet delivery ratio with aggregation of data packets. [4]

II. Aggregation Function and Energy Calculations

Abbreviations: K=packets generated at different PGR.
M=packets generated at equal PGR.
X=sum of packets due to different PGR.
Y=average of the number of packets generated by equal PGR.

If Xi and Yj are two variables that represents the number of packets generated (different and Equal) by the 'u' and 'h' participating nodes in the cluster, provided that i=1...K and j=1...M, then the perfectly compressible aggregation function will be Additive functions: If each of the heterogeneous node in the specified region of the cluster has variable PGR then the aggregation function is[5],

$$\{sum\text{-}f(A_s)\text{-}\sum_{i\text{-}1}(Xi) \text{ for all } (X_i)\text{- different PGR}$$

$$\{X_i\text{-}(p_{r1}+p_{r2}+p_{r3}+\ldots\ldots.+p) \text{ //sum of packets by nodes} \quad (1)$$

Divisible aggregation function: If PGR of the entire heterogeneous node is same then the aggregation function is,
Average-$f(A_v)$-$1/m\sum^{m}_{j\text{-}i}(Y_j)$ for all(Y_j)-equal PGR

$$Y_j\text{-}(\, p_{r1}+p_{r2}+p_{r3}+\ldots\ldots.+p_{ru}\,)/_{nu,h} \quad (2)$$

The total initial energy in the network is With the consideration of heterogeneity of nodes in the network, the total initial energy of network is;

$$E_i\text{-}N\,(\alpha E_n+\beta E_a+\gamma E_s) \quad (3)$$

The cost of aggregation according to is

$$C_i\text{-}\sum_{j\varepsilon Ni} \cos t(n,CH) + \cos\ t(\,CH,\sin k) \quad (4)$$

This indicates the energy consumed by a cluster member 'n' to send a packet to CH and CH to sink.

III. PROPOSED APPROACH

3.1 MHCDA ALGORITHM

The basic design objectives of the MHCDA for aggregation and bandwidth utilization are:
•Elect the CH and decide the aggregation function.
•Minimize the communication cost with better bandwidth utilization, increase the network lifetime with reduced energy consumption.
•Analyze the effect of mobility of sink on the performance of the algorithm.

The MHCDA algorithm for improving bandwidth utilization and a decrease in energy consumption is divided into three phases as cluster formation, intra-cluster aggregation, and inter-cluster aggregation as shown in Fig.2. In the first phase, all the heterogeneous nodes distribution is random and becomes a member of any one cluster. one CH is elected according to the highest energy among the nodes (normal, advanced, or super) and the highest number of average neighbor nodes. Re-election of CH has not been considered, thus, saving energy. In the second phase, CH is responsible for aggregation of the data packets of a fixed size. In the third phase, the aggregated packets from CH are forwarded to the mobile sink with inter-cluster aggregation at multi-hop level. This reduces the communication cost.

To perform aggregation, the CH takes the spatial and temporal relations into account in terms of PGR of source nodes and random data within cluster limits. This phase runs recursively for all the clusters within the network for intra-cluster aggregation. In the third phase, each CH is considered as one node for performing inter-cluster aggregation and communicates the aggregated packets/data to the sink by forming graph of CH as G={C1,C2,C3,...,Cn}. The aggregation function applied on packets at intra and inter cluster aggregation reduces the packet count hence PDR and throughput. The algorithm considers the effect of packet and data aggregation of MHCDA with and without them mobility of sink.

3.1.1 DETAILS OF ALGORITHM

This section deals with the actual working phases of MHCDA. MHCDA works in three phases as cluster formation, intra-cluster and inter-cluster aggregation [6].

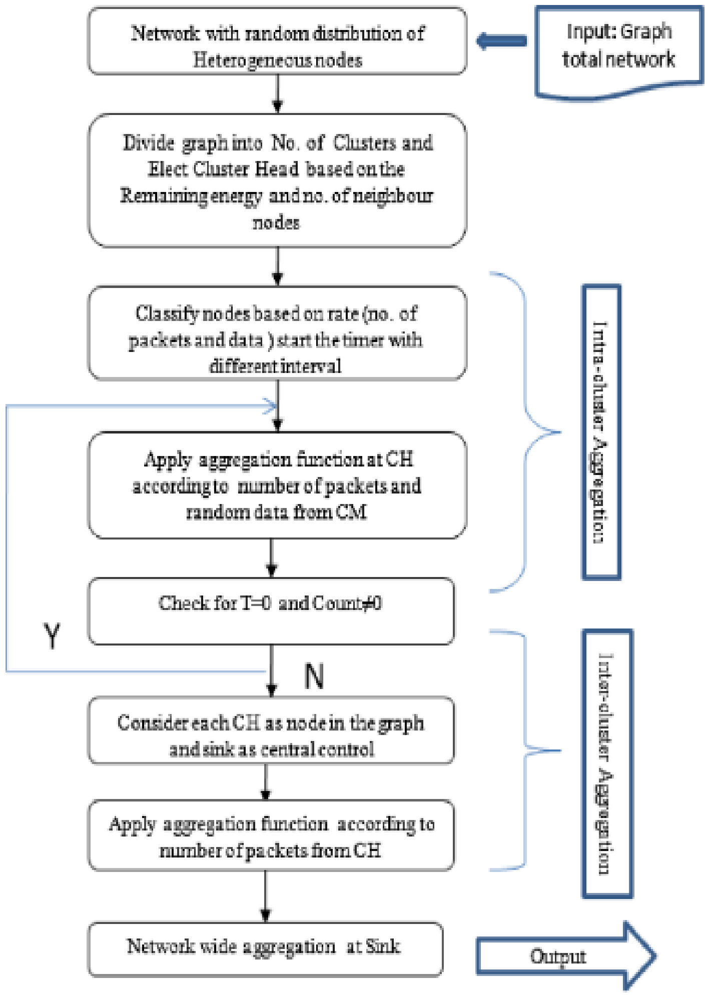

Fig2. Flow Chart for the step to be followed

3.1.2 INTRA-CLUSTER AGGREGATION

It is assumed that nodes are heterogeneous and part of the network graph G with 'N' number of nodes into each cluster. Let P denote the set of all packets received by the CH from the nodes within the region R. CH performs the aggregation on the packets arrived at t=0 and calculates effective count Cp>0. Let fp(t) denote the number of packets aggregated at time t>0. The state of CH at time t is denoted by its vector f(t)={fp(t)}.Note that each node u, h ε R will transmit a packet to CH with a set of active links (u, h,CH). Let 'Ps' denote the set of all possible packets received, andP(u,h) denote the set of packets transmitted when node 'u,h' are active. Therefore, an intra-cluster aggregation function based on the current number of packets from nodes at time t calculated according to Eqs.1 and 2 is.

$$f(Ps,CH)\text{-}\sum p\varepsilon PCp\{p(u,h)\epsilon R,fp(t)\} \qquad (5)$$

Algorithm 1 gives details of CH formation and intra-cluster aggregation.
Input:Graph G (V, E) with 'n' clusters distributed in ClustersOutput: CH with aggregated information

```
1. Sink node generate parent announcement message (Par_ann_msg)
2. Set Par_ann_msg -> parent_ch = Sink_id
     Par_ann_msg -> hop_count =1
     Par_ann_msg ->Previous_hop=Sink_id
3. Initialize the timer for periodically broadcast Par_ann_msg
4. If Par_ann_msg is received by CM
   If (Routing_table (seq_no) < Par_ann_msg ->seq_no)
   {
     Update the Routing Table
     (Next_hop) = Par_ann_msg – >Previous_hop:
     Hop_count++,Previous_hop->current node id
     Par_ann_msg ->hop count++;
```

```
    Par_ann_msg ->Previous_hop=current_node_id
    Forward to CH }
Else
 {If (Routing_table (seq_no) == Par_ann_msg ->seq_no)
 {If(Routing_table(hop_count)<Par_ann_msg->hop_count) {
Routing_table(Next_hop)=Par_ann_msg>Previous_hop
  Increment hop count
  Par_ann_msg->hopcount++;
  Par_ann_msg ->Previous_hop=current_node_id
  Forward to CH } } }
Else
 {Drop the Packet }
5. If Par_ann_msg is received by CH
    If (Routing_table (seq_no) < Par_ann_msg ->seq_no)
 { parent_ch = Par_ann_msg -> parent_ch;
   (Next_hop) = Par_ann_msg –> Previous _hop:
   change parent cluster id to its node id
   Par_ann_msg -> parent_ch=current_node_id;
 Forward to other CH }
Else
{ If (Routing_table (seq_no) == Par_ann_msg ->seq_no){
   If(Routing_table(hop_count)<Par_ann_msg–>hop_count)    Sets parent_ch = Par_ann_msg -> parent_ch ,
   Routing_table(Next_hop)=Par_ann_msg>Previous_hop
   Par_ann_msg -> parent_ch=current_node_id;
} } }
Else
{Drop the Packet}
End if
```

3.1.3 Inter-cluster Aggregation

In the third phase, graph G includes the sink and all participating CH as G={CH1,CH2,...CHn} with all V {CH, sink} and E care the connecting edges between all the cluster heads and the sink. The algorithm 2 used for the intra-cluster aggregation operates recursively for the inter-cluster aggregation with the same assumptions.

Algorithm2: for Inter-Cluster Aggregation

```
1.    All the CH acts as nodes with aggregated packets
          {If (intra-cluster aggregation) then
           If (event of interest)  then
           Generate packets and data at a variable rate
            Forward to CH
Apply the aggregation function at CH according to equation 1 and 2
else
Store it and aggregate with other readings
End if
Else
Generate the data and packets using a variable rate and a random
Function
End if
 Forward to CH (Aggregator)}
2.    Collection Phase:
{
If (packet reaches to the aggregator)
Store into routing table
   If (Previous packet/ data= next Packet/ data) then
     Drop the packet/data (no aggregation)
Else
     Wait for T sec/count
        If (count =0) then
          Apply perfectly compressible aggregation function
End if
End if
End if
}
3. Provide mobility to sink
```

4. Sink with all the aggregated packets.

IV.BHCDA ALGORITHM

4.1.Network Model

The network model is the connecting graph G (V, E) of different clusters {C_1, C_2, C_3, - - - C_n} of WSNs with heterogeneous nodes and mobile sink. Each region has a CH and number of nodes that are represented by a set of vertices 'V' and wireless connecting edges 'E'. The 'V' nodes in the network randomly distribute and organize into 'n' clusters using a multi-hop clustering algorithm. Some of the nodes 'h' are deployed with higher energy (30J, 40J) than the normal nodes 'u' (20J). Now consider that each cluster has 'N' nodes out of which 'u' and 'h' nodes ' u, h ὁ N' acts as cluster member and generates the variable data packets of fixed size[7]. To aggregate the data packets the perfectly compressible aggregation functions at CH is defined as:

$$f(C_A)=\sum^{k}_{i=1}(X_i)+1/M\sum^{M}_{j=1}(Y_j) \quad (1)$$

Where X_i and Y_j are variables that represent the correlation of the number of data packets generated by the 'u' and 'h' participating nodes in the cluster.

The aggregation functions used in BHCDA aims to

• Increase the bandwidth utilization (intra-cluster aggregation).
•Reduce the communication overhead (cost).
•Minimize energy consumption in the network.

$$E_i=N(\alpha E_n+\beta E_a+\gamma E_s) \quad (2)$$

4.1.1 Details of Algorithm

The BHCDA is presented for improving bandwidth utilization in terms of the packets delivery ratio and throughput as metric of computation with decrease in energy consumption. It works with three phases as show in Figure 3.In the initial phase, randomly distributed heterogeneous nodes are organized into number of clusters. It decides the CH for intra-cluster and inter-cluster aggregation from each square region. It is elected according to the highest energy among the cluster members (normal, advanced or super node) highest number of average neighbor nodes with one hop connectivity. The threshold for the node to become CH is revised as:

$$T(n)=1/1\text{-}p(r\bmod 1/p)\ E_r/E_i\ D_i/D_{avg} \quad (3)$$

Where 'P' is the percentage probability of selecting cluster head, 'r' is the current round,(r mod 1/p) is the representation of elected cluster heads. 'E_r' is residual energy after each round and 'Ei' is the initial energy. 'D_i' is the degree of connectivity of the '$^{\delta}$'is the density of node within the selected range of cluster before the election of cluster head. Residual energy, density of node and connectivity decides the election of cluster head, once elected it continues.

In the Second Phase, each CH responsible for aggregation of data packets generated by cluster members with variable rate. The size of data packet generated is fixed.

According to the network consideration, each cluster head in the third phase works like individual node and perform inter-cluster aggregation. The final result is the aggregation of data packets at the sink with reduced packet count, hence reduced transmission cost. Let C_i denotes the cluster formed by CH node 'i' with its neighbor nodes 'Ni', then the cost of aggregation with cluster C_i is defined according to[8] is

$$C_{i=\sum j\epsilon Ni}\ COS\ t(j,i)+COS\ t(i,\sin K) \quad (4)$$

This indicates the cost of data aggregation in terms of energy consumption by the network,where cluster members sends random packets to Ch [cost (j,i)] and CH to sink {cot(i,sink)}

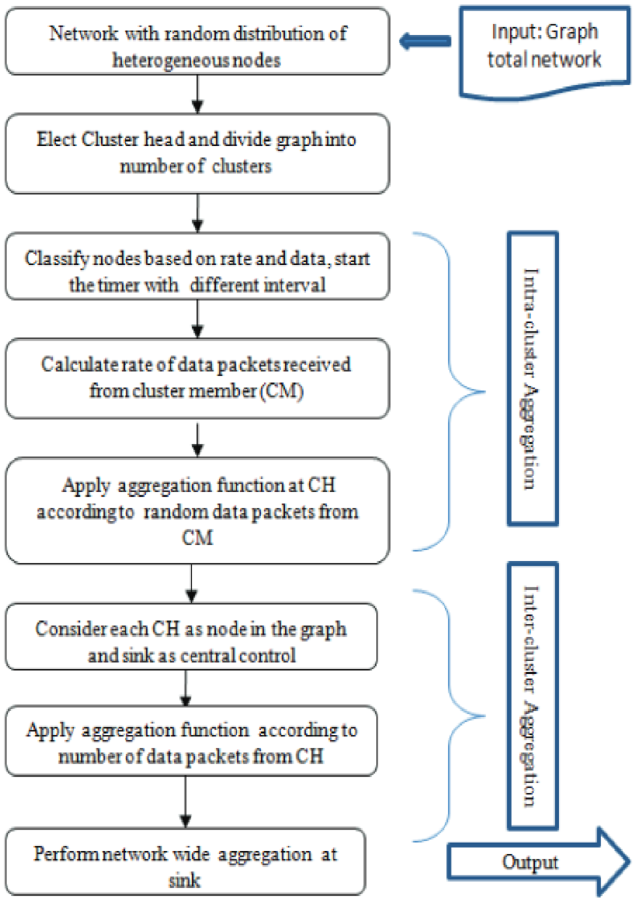

Fig3. Flow Chart for the step to be followed

With the assumption that nodes in the networks have equal density and heterogeneous in energy level are the parts of the networks graph G.The aggregated information for the intra cluster phase with nodes 'N' in to the cluster is obtained by the following algorithm.

Input: Graph G(V,E) with 'n' cluster G={C1,C2,C3,-Cn}
Ouput: CH with aggregated data packets and sink with network wide aggregation.

```
1.Sink node generates parent announcement message
        (par_ann_msg)
2.Set Par_ann_msg->parent_ch=Sink_id
   Par_ann_msg->hop_cont=1
   Par_ann_msg->Previous_hop=sink_id
3.Initialize the timer for periodically broadcast    Par_ann_msg
    If Par_ann_ms is received by CM
    If (Routing_table(seq_no)<Par_ann_msg->seq_no)
    {
       Update the Routing table
        Routing_table(Next_hop)=Par_ann_msg->Previous_hop
       Increment Hop count and update previous hop as current node id
       Par_ann_msg->hop cont++;
       Par_ann_msg->Previous_hop=current_node_id
       Forward to CH
          }
          Else
         {
           If(Routing_table(seq_no)==Par_ann_msg->seq_no)
            {
                  If(Routing_table (hop_count)<Par_ann_msg->
                  hop_count):update the Routing Table
                  Routing_table(Next_hop)=Par_ann_msg->
                  Previous_hop: Increment hop count
                  Par_ann_msg->hop cont++;
                  Par_ann_msg->Pervious_hop=current_node_id
                 Forward to CH
      }
}
```

```
{

  Drop the packet
}
4.For Data forwarding to CH and SINK

{ If(intra cluster aggregation) then
   If (event of interest), then
     Generate data at a variable rate
      Forward to CH
Else
     Store it and aggregate with other readings
End if
Else
      Generate the data packets
end if
      generate the data packets
end if
Forward to CH (Aggregator)
}
Collection Phase:
{
        If(packet reaches to aggregator)
       Store in to the routing table
       If(Previous data packets packet data =next data packet data)then
       Drop the data packet( no aggregation)
Else,
       comparison
       Wait for T sec/count
        If (count=0) then,
       apply aggregation function
end if
end if
end if
```

V.CONCLUSIONS

The proposed cluster-based data aggregation algorithm MHCDA is applicable to IoT applications using mobility as one of the parameter. The proposed algorithm applies the additive and divisible aggregation functions on the packets and data generated at variable rate by considering the spatial and temporal correlation of packet and random data. From the results, it is observed that the packet count reached at the sink reduces with an improvement in communication cost and proposed novel techniques to improve the network performance by aggregating the variable rate of data packets generated by heterogeneous nodes in the network with mobile sink. It uses the perfectly compressible aggregation function to increase the bandwidth utilization and energy saving.

REFERENCES

[1] Mainetti, L., Patrono, L., & Vilei, A. (2011). Evolution of wireless sensor networks towards the internetof things: A survey. In19th international conference on software, telecommunications and computer networks (SoftCOM) (pp. 1–6).

[2] Wang, L., Da Xu, L., Bi, Z., & Xu, Y. (2014). Data Cleaning for RFID and WSN Integration.IEEETransactions on Industrial Informatics, 10(1), 408–418.

[3] Dnyaneshwar S.Mantri.Neeli Rashmi Prasad.Ramjee Prasad (2015). Mobility and Heterogeneity Aware Cluster-based Data Aggregation for Wireless Sensor Networks.Wireless pers commun(2016)86:975-993.

[4] Dnyaneshwar Mantri,Neeli Rashmi Prasad,Ramjee Prasad (2013) Bandwidth Efficient Heterogeneity aware cluater based Data Aggregation for Wireless Sensor Network.IEEE International Conference On Advances in Computing,Communications and Informatics.

[5] Kumar, D., Aseri, T. C., & Patel, R. B. (2011). EECDA: Energy efficient clustering and data aggre-gation protocol for heterogeneous wireless sensor networks.International Journal of ComputersCommunications and Control, 6(1), 113–124. ISSN 1841-9836.

[6] Khan, M. I., Gansterer, W. N, & Haring, G. (2012). Static vs. mobile sink: The influence of basicparameters on energy efficiency in wireless sensor networks.Computer Communications. ISSN0140-3664.

[7] Y. Mao, Z. Liu, L. Zhang, and X. Li, "An Eœective Data Gathering Scheme in Heterogeneous Energy Wireless Sensor Networks," Proceedings of International Conference on Computational Science and Engineering, pp.338-343, 2009.

[8] Li Han,"An energy efficient routing algorithm for heterogeneous wireless sensor networks," 2010 International Conference on Computer Application and System Modeling (ICCASM), vol.3, pp.612-V3-616, 22-24 Oct. 2010