CHAPTER 4

CLUSTER BASED COMPRESSED DATA AGGREGATION
AND ROUTING IN WSN

4.1 OVERVIEW

In existing cluster based data aggregation approaches, the load at the CH will be more leading to the quick energy drain of the CH, since all the members transmit their sensed data to the optimal cluster head. Moreover, due to redundancy in data, the aggregation overhead will be more. The inter-cluster communication should ensure failure free transmission and energy efficiency. Hence the objective of this works are reducing the data redundancy, minimizing the load and energy drain rate of the cluster head and ensuring energy efficient inter-cluster communication in cluster based data aggregation.

To meet these objectives, a data correlation based compression model and relay based inter-cluster routing algorithm are proposed and applied in the cluster based WSN. This is an extension to the optimal clustering protocol discussed in chapter-3.

Here, data correlation and data compression method is used as an energy saving method. With this, the compression ratio is calculated from the data correlation and the cluster size. The Cluster Head (CH) aggregates and compresses the collected readings and transmits a single representative message to the sink. The size of the compressed message depends on the joint entropy of the cluster. The following Figure 4.1 shows the block diagram of proposed work.
4.2 OPTIMAL CLUSTERING PROTOCOL

1. At time t, several provisional cluster heads (PCH) are elected in a random manner to compete for final cluster heads.

2. Excluding the chosen PCH, the remaining nodes broadcasts a HELLO message at fixed power level.

3. Each PCH which receives HELLO message estimates the distance and appends the node ID to its route cache.

4. Finally, for each PCH, the parameters such as Residual Energy (Eres), Node Density (NDi) and Load are gathered.
5. Utilizing the estimated parameter values, each node analyzes the parameters condition using fuzzy logic technique. Fuzzy logic handles the uncertainties in electing cluster heads and determining the cluster size.

6. When a PCH estimates its possibility to become the cluster head, broadcasts a desire message (D_Mes) establishing its desire to become CH.

7. However, when PCH finds that there is another PCH with the greater criteria to get selected as CH, it just declaration its desire cancellation (DCL_Mes) message.

8. After CH election, each CH broadcasts a cluster advertisement (CL_ADV) message through the network. Ordinary sensors nodes in the network join the nearby CH.

4.3 DATA CORRELATION MODEL

The sensor nodes within the WSN are considered statistically identical information sources with an assumption that the readings are normally distributed with mean zero and variance. In case of degree of data correlation (DDC), the assumption is dependency between sample readings exponentially decreases with distance at a fixed rate. So if the distance is more, the correlation will be less.

Hence, in order to evaluate the DDC of the two time series, the following two metrics magnitude and distance are defined.

The readings of two sensor nodes $N_i$ and $N_j$ are said to be correlated, if

(i) $MG_i = MG_j$

where $MG_i$ and $MG_j$ are the magnitudes of the values of $N_i$ and $N_j$
(ii) \( D_{ij} < D_{th} \)

where \( D_{ij} \) is the distance between \( N_i \) and \( N_j \) and \( D_{th} \) is the distance threshold

The two nodes are represented as 2- dimensional points. Node \( N_i \) has coordinate \((M_{Gi}, D_i)\) and Node \( N_j \) has coordinate \((M_{Gj}, D_j)\). The DDC between the two nodes is based on these coordinates estimated using the Euclidean Distance is given in 4.1.

Euclidean distance examines the root of square differences between coordinates of a pair of objects. Therefore, the Euclidean Distance between the nodes \( N_i \) and \( N_j \) is given by,

\[
\text{DDC}_i = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}, \quad n = 3.
\] (4.1)

here, \( x_{i1} = M_{Gi}, \ x_{i2} = D_i \) and \( x_{j1} = M_{Gj}, \ x_{j2} = D_j \).

This data correlation model is used to allocate the optimal set of data rates to the sensor nodes. Here the Power Exponential model is used as a covariance function for this data correlation model.

The Power Exponential model for the covariance function with the distance \( d \) is given by Equation 4.2:

\[
K_{\mu}^{PE}(d) = \exp\left(-\frac{d}{t_1}\right),
\] (4.2)

Here, \( K_{\mu}^{PE}(\|u - v\|) = corr\{S(u), S(v)\} \) denotes an isotropic correlation function with \( \mu = (t_1, \ldots, t_c) \in \Theta \subset \mathbb{R}^c \) as set of parameters, where \( t_1 > 0 \) and \( t_2 \in (0, 2] \).
4.4 DATA COMPRESSION MODEL

Using the data compression model, the size of the compressed message depends upon the joint entropy. Here the correlated data streams are encoded independently, and then decoded jointly at one receiver.

The data streams in the nodes $N_i$ and $N_j$ are said to be correlated, such that the data stream in node $N_i$ is $X$ and the data stream in node $N_j$ is $Y$.

The correlated data streams $X$ and $Y$ is generated by making $n$ independent drawings from a joint probability distribution $P(X = x, Y = y)$.

Figure 4.2 Independent Encoding and Joint Decoding of two correlated data streams $X$ and $Y$

In Figure 4.2, Encoder 1 receives data stream $X$ and then transmits a coded message to the decoder, each character of $X$ is encoded by the $R_X$ bits. Likewise, encoder 2 receives data stream $Y$ and then transmits a coded message to the decoder, where each character of $Y$ is encoded by a number of $R_Y$ bits. On receiving these two coded messages, the decoder will generate two $n$-vectors $X^*$ and $Y^*$, which represents the estimations of the original data streams $X$ and $Y$, respectively.
The admissible rate region for the random variables \( X \) and \( Y \) with joint probability distribution \( P(X = x, Y = y) \) is given by Equation 4.3,

\[
H(X \mid Y) = -\sum_y P(Y = y) \sum_x P(X = x \mid Y = y) \log P(X = x \mid Y = y),
\]  
(4.3)

By the Slepian-Wolf theorem, the admissible rate region for the pair of rates \((R_X, R_Y)\) satisfies the following three inequalities:

- \( R_X \geq H(X \mid Y) \)
- \( R_Y \geq H(Y \mid X) \)
- \( R_X + R_Y \geq H(X, Y) \)

The separate encoders that ignore the source correlation can achieve rates if only \( R_X + R_Y \geq H(X, Y) \).

4.5 INTER-CLUSTER ROUTING ALGORITHM

Each cluster head broadcast the routing message with the information’s such as the id, the residual energy, the number of cluster members and the distance to the BS of itself. The Probability of next hop CH is given by Equation 4.4.

\[
PR_{CHN} = \alpha \frac{E_j}{E_{max}} + (1 - \alpha) \frac{1}{(NO\_CM)_j}
\]  
(4.4)

Where
- \( E_j \) is the residual energy of cluster head \( S_j \)
- \( E_{max} \) is the maximum initial energy of nodes
- \((NO\_CM)_j \) is the number of cluster members in cluster head \( S_j \)
- \( \alpha \) is a constant whose values lies between \([0,1]\).
**Notations**

RT_MSG is the route message which consists of id (Id), residual energy (RE), number of cluster members (No_CM) and distance to BS (Dist_BS).

DIST_TH - threshold distance

T_c - Timer for constructing the routing tree

CHNT – Cluster head neighbor table

PR_CHN – Probability of next hop CH

The pictorial representation is given in Figure 4.3 and its inter-cluster routing algorithm is given below.

![Inter-cluster routing diagram](figure43.png)

**Figure 4.3** Representation of inter-cluster routing with the encoded data packets
Algorithm

1. CH broadcast a RT_MSG

   CH \rightarrow \mbox{RT\_MSG [Id, RE, No\_CM, Dist\_BS]}

2. If \quad \mbox{Dist\_BS < DIST\_TH, then}

   Data can be transmitted directly to BS

   Else

   While (timer Tc not expired)

   \quad Compute the value of PR_{CHN} for each neighbor CH.

   \quad Update CHNT

   End

   For each CH\_j in CHNT

   \quad If PR_{CHN} = \mbox{Maximum}(PR_{CHN}), then

   \quad \quad Choose CH\_j as the next hop CH.

   \quad \quad End if

   End for

   If j > 1, then

   \quad \quad Choose CH\_j with less Dist\_BS

   \quad \quad End if

   End if

The following Figure 4.4 shows the inter-cluster routing algorithm flow diagram.
CH broadcasts a Route_Msg

\[
\text{disttoBS} < \text{DIST\_TH}
\]

Yes \rightarrow Next hop BS is selected

No \rightarrow while (T_4 is not expired) Receive Route_Msg

Compute the value of relay

Update CH neighborhood table CHNT

If CH \( s_m \) has no cluster member

Yes \rightarrow Next hop \( s_m \) is selected

No \rightarrow \( s_j \) has the max value of relay in CHNT

Yes \rightarrow Update Max value of relay (MR)

No \rightarrow \( s_k \) has the max value of disttoBS in MR

Next hop \( s_k \) is selected

End

Figure 4.4 Inter-Cluster Routing Flow Diagram
4.6 SIMULATION RESULTS

The Network Simulator (Network Simulator: http://www.isi.edu/nsnam/ns) is used to simulate the proposed architecture.

4.6.1 Simulation Model and Parameters

The following table 4.1 shows the simulation settings and parameters which are summarized as,

<table>
<thead>
<tr>
<th>Table 4.1 Simulation Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of Nodes</strong></td>
</tr>
<tr>
<td><strong>Area Size</strong></td>
</tr>
<tr>
<td><strong>Mac</strong></td>
</tr>
<tr>
<td><strong>Transmission Range</strong></td>
</tr>
<tr>
<td><strong>Simulation Time</strong></td>
</tr>
<tr>
<td><strong>Traffic Source</strong></td>
</tr>
<tr>
<td><strong>Packet Size</strong></td>
</tr>
<tr>
<td><strong>Initial Energy</strong></td>
</tr>
<tr>
<td><strong>Receiving Power</strong></td>
</tr>
<tr>
<td><strong>Transmission Power</strong></td>
</tr>
<tr>
<td><strong>Rate</strong></td>
</tr>
</tbody>
</table>
4.6.2 Performance Metrics

The proposed Cluster based Compressed Data Aggregation and Routing (CCDAR) protocol is compared with the Cluster based Correlated Data Gathering (CCDG) technique (Ali Dabirmoghaddam et al., 2010). The performance metrics Packet Delivery Ratio and average Energy Consumption are evaluated.

- **Packet Delivery Ratio**: It is the ratio between the number of packets received and the number of packets sent.

- **Packet Drop**: It refers the average number of packets dropped during the transmission.

- **Energy Consumption**: It is the amount of energy consumed by the nodes to transmit the data packets to the receiver.

- **Delay**: It is the amount of time taken by the nodes to transmit the data packets.

4.6.3 Results

The performance of CCDAR and CCDG is evaluated by varying the nodes as 25, 50, 75 and 100. Table 4.2 presents the results of both the techniques for varying the nodes.

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Delay</th>
<th>Delivery Ratio</th>
<th>Drop</th>
<th>Energy</th>
<th>Overhead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCDAR</td>
<td>CCDG</td>
<td>CCDAR</td>
<td>CCDG</td>
<td>CCDAR</td>
</tr>
<tr>
<td>25</td>
<td>12.28903</td>
<td>12.40088</td>
<td>0.599847</td>
<td>0.568029</td>
<td>4687</td>
</tr>
<tr>
<td>50</td>
<td>7.262955</td>
<td>8.869068</td>
<td>0.492422</td>
<td>0.425272</td>
<td>5573</td>
</tr>
<tr>
<td>75</td>
<td>6.144103</td>
<td>9.778721</td>
<td>0.375955</td>
<td>0.280947</td>
<td>6306</td>
</tr>
<tr>
<td>100</td>
<td>7.442191</td>
<td>10.69368</td>
<td>0.369848</td>
<td>0.269417</td>
<td>6500</td>
</tr>
</tbody>
</table>
Figure 4.5 Comparison based on Nodes Vs Delay

Figure 4.5 shows the results of delay for CCDAR and CCDG techniques, when the number of nodes is increased. As we can see from the figure, the delay starts to increase from 50 nodes for both the techniques. However, since CCDAR involves inter-cluster routing, the delay is 21% less, when compared to CCDG.

Figure 4.6 Comparison based on Nodes Vs Delivery Ratio
Figure 4.6 shows the results of packet delivery ratio for CCDAR and CCDG techniques, when the number of nodes is increased. As we can see from the figure, the delivery ratio decreases from 0.55 to 0.36 for CCDAR and 0.56 to 0.26 for CCDG. However, since CCDAR involves relay nodes for inter-cluster routing, it attains 17% higher delivery ratio when compared to CCDG.

![Nodes Vs Packet Drop](image)

**Figure 4.7 Comparison based on Nodes Vs Packet Drop**

Figure 4.7 shows the results of packet drop occurred for CCDAR and CCDG techniques, when the number of nodes is increased. As we can see from the figure, the packet drop increases from 4687 to 6500 for CCDAR and 6917 to 13,377 for CCDG. Since CCDAR involves relay nodes for inter-cluster routing, the packet drop is reduced by 44% when compared to CCDG.
Figure 4.8 Comparison based on Nodes Vs Energy Consumption

Figure 4.8 shows the results of average energy consumption for CCDAR and CCDG techniques, when the number of nodes is increased. As we can see from the figure, the energy consumption remains constant beyond 50 nodes for both the techniques. Since CCDAR applies compressed data aggregation along with energy efficient inter-cluster routing, the energy consumption is reduced by 11% when compared to CCDG.

Figure 4.9 Comparison based on Nodes Vs Overhead
Figure 4.9 shows the results of average control overhead occurred for CCDAR and CCDG techniques, when the number of nodes is increased. As we can see from the figure, the overhead remains constant beyond 50 nodes for both the techniques. Since CCDAR applies compressed data aggregation along with energy efficient inter-cluster routing, the overhead is reduced by 47% when compared to CCDG.

Table 4.3 presents the percentage wise improvement of CCDAR over CCDG by varying the number of nodes.

**Table 4.3 Percentage wise improvement of CCDAR by varying the number of nodes**

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Delay (%)</th>
<th>Delivery Ratio (%)</th>
<th>Drop (%)</th>
<th>Energy (%)</th>
<th>Overhead (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.901968</td>
<td>5.304353</td>
<td>32.23941</td>
<td>2.545409</td>
<td>42.60639</td>
</tr>
<tr>
<td>50</td>
<td>18.10915</td>
<td>13.63668</td>
<td>43.08038</td>
<td>14.639</td>
<td>49.61488</td>
</tr>
<tr>
<td>75</td>
<td>37.16864</td>
<td>25.27111</td>
<td>50.76899</td>
<td>12.26673</td>
<td>50.12191</td>
</tr>
<tr>
<td>100</td>
<td>30.40569</td>
<td>5.240761</td>
<td>46.99817</td>
<td>14.85982</td>
<td>48.63778</td>
</tr>
</tbody>
</table>

**4.7 CONCLUSION**

In this chapter, an efficient data aggregation tree has been proposed for communication and routing based on the previous clustering architecture. Here, data correlation and data compression method is used as an energy saving method. Next, a routing tree is constructed using the cluster based routing algorithm. This cluster-based inter-cluster routing algorithm balances the energy consumption among cluster heads by adjusting intra-cluster energy consumption and inter-cluster energy consumption. By simulation results, it is shown that the proposed technique enhances the network lifetime with increased delivery ratio and decreased energy consumption. In the next chapter, we will discuss about the proposed energy efficient sleep-scheduling for cluster based aggregation in WSN in which scheduling is performed with efficient energy consumption.