CHAPTER 4

CLUSTER LOCALIZATION OF SENSOR NODES USING
HIDDEN MARKOV MODEL

4.1 INTRODUCTION

This chapter presents an idea to localize the sensor nodes through clustering. Clustering is a standard approach for achieving efficient and scalable performance in WSN. To achieve the network scalability, sensor nodes are grouped to form clusters. The advantage of clustering is that it can localize the route set up within the cluster. Abbasi & Younis (2007) have surveyed on clustering algorithms for WSN and shown the feature of LEACH does not need the location awareness based on the clustering attributes. Cluster saves energy and reduces network contention through the cluster head (CH). The idea of LEACH algorithm is to form clusters of sensor nodes based on RSS, and use local CH as routers to base station (Heinzelman et al 2000). LEACH saves energy since the transmissions are mainly managed by CH. Initially CH are randomly selected and changed over time in order to spread load and balance the energy. LEACH assumes that each node has enough power to transmit signals to reach CH. This is not applicable to deploy in large regions due to the variation of distances between sensors and CH (Akyildiz et al 2002).

Moreover, the idea of dynamic clustering brings extra overhead, such as rotation of CH and accordingly consumes energy. The disadvantage of LEACH is the CH is elected randomly, so the optimal number and
distribution of CH cannot be ensured. The node with low energy has the same priority to be a CH as the node with high energy. Therefore, those nodes with less remaining energy may be chosen as the CH wherein these nodes may die first. The cluster head communicate with the base station in single-hop mode and hence LEACH cannot be used in large-scale wireless sensor networks for the limited effective communication range of the sensor nodes. These disadvantages of LEACH are overcome by proposing a new Coverage Preserving Clustering Protocol (CPCP). Cluster localization is achieved by keeping the CH static and the cluster members dynamic. The location of cluster members is obtained with the help of CH that has the knowledge of its location. CPCP approach is proposed to form clusters for the localization with HMM and is compared with existing LEACH protocol.

4.2 PROPOSED WORK

The main objective of the proposed work is to attain cluster localization through HMM. CPCP is used for CH election and cluster formation. The sensor activation phase of CPCP supports to check that all nodes in the network are covered to form clusters. After all nodes in the network forms into clusters, CH is elected and assumed to be static. The cluster members in the cluster moving dynamically are localized using HMM. The nodes follow the movement models and performance comparison using simulation is done with existing LEACH for HMM.

4.3 MOTIVATION

Mengual et al (2010) proposed a three-phase methodology (measurement, calibration and estimation) for locating Mobile Stations (MS) in an indoor environment using wireless technology. RSS measurements collected at different locations uses Artificial Neural Network models (ANNs) to group them into clusters. Clustering using data mining technique is used to
4.4 SYSTEM DESIGN

The system design includes three modules namely Network model, Formation of Clusters and Location estimation using clustering protocol with HMM. Each module is described in detail in the subsequent sections.

4.4.1 Network Model

Assume the network in an outdoor region of sensor nodes $S_N$ from a set $S$, $s_i \in S$, where $i = 1, \ldots, S_N$ is deployed randomly over a monitored area. Each sensor performs reliable sensing within its sensing area $C(s_i)$, which is approximated by a circular area around the node with radius $R_{\text{sense}}$. The area is divided into small grids of size $n \times n$ wherein the anchor node (CH) is spread randomly among regular sensor nodes that have the knowledge of its location from the cluster head election. Other non-anchor sensor nodes (cluster members) estimate their location using RSS based on the location of anchors. The CH is assumed to be static and the cluster member is moving dynamically over the network. The cluster is constructed with the help of randomly deployed nodes to form the cluster. The cluster radius $R_{\text{cluster}}$ is a tunable parameter that determines the minimum distance between any two CH nodes in the network and is shown in Figure 4.1. This parameter can be easily tuned by changing the transmission power of the CH nodes.
4.4.2 Formation of Clusters

Soro & Heinzelman (2009) has formed clusters through CPCP. CPCP consists of six phases such as Information update phase, CH election phase, Route update phase, Cluster formation phase, Sensor activation phase and Data communication phase for coverage preservation.

4.4.2.1 Information update phase

This phase updates information on the remaining energies of the nodes. Each sensor node broadcasts an update packet with information about its remaining energy to all its neighbors in the range $R_{\text{cluster}} = 2 \times R_{\text{sense}}$. If energy-aware cost $C_{\text{ea}}$ is used, then this Information Update phase can be skipped, since these cost metrics do not depend on the neighboring nodes’ remaining energies. Energy-aware cost $C_{\text{ea}} = 1/E(s)$ where $E(s)$ is the remaining energy.
4.4.2.2 Cluster head election phase

In this phase, every sensor governs its “activation time”, i.e., an amount of time proportional to its current energy-aware cost $C_{ea}$. The $C_{ea}$ is the sensor’s ability to take part in the sensing task based on its remaining energy. Each sensor has to wait for the expiration of its activation time before deciding whether or not the sensor should announce itself as a new CH. During the activation time, if a node does not hear an announcement message from any other sensor node, then, after expiration of its activation time it declares itself to be a new CH, by sending an announcement message to all the nodes within the cluster range. The announcement message contains information about the new CH location.

4.4.2.3 Route update phase

All cluster heads from where it receives the announcement message so far as well as the distance to each CH node is updated and maintained in a table. In this manner, the CH election takes place, moves to form the cluster and the routes are updated.

4.4.2.4 Cluster formation phase

In the cluster formation phase, each non-CH node decides to join the closest CH node in order to form the cluster. By sending the JOIN messages to the selected CH nodes, a node becomes a member of the cluster for the upcoming round. Upon this activation message cluster uses low energy for the cluster formation. There are no restrictions on the number of cluster members in the cluster. Based on cluster range the nodes join less than one CH in the network. In this way, the cluster formation takes place around CH nodes in the network to enhance lifetime of the network.
4.4.2.5 Sensor activation phase

In the sensor activation phase, each sensor node assigns itself an activation delay that is proportional to its current energy-aware cost $C_{ea}$. In this way, sensor nodes with smaller cost waits for a period of time before deciding whether it has to stay awake during the next communication round. While waiting for its activation delay to expire, the sensor node can receive the ACTIVATION messages from its neighboring nodes. It can have smaller activation delays (smaller cost), if they decide to become active during the upcoming communication round. After its activation delay time expires, the sensor node determines its sensing area is completely covered by its neighboring nodes; it turns itself off for the upcoming round. Otherwise, the sensor node broadcasts an ACTIVATION message to inform its neighbors about its decision to remain active. In this way, the lower cost nodes have priority to decide whether they should be active.

The sensor range in the sensor activation phase is obtained by initializing the sensor nodes $S_N$ to $E(s)$, where $E(s)$ is the residual energy of node $S_N$. It checks the sensor activation time $T_a(s)$ that is proportional to sensing area where it belongs to energy-aware cost of the sensor node. During the period $T_a(s)$, $S_N$ can receive ACTIVATION message from its neighboring nodes. Once the $T_a(s)$ expires, $S_N$ checks whether the sensing range is fully covered. If the sensing range of node is not fully covered then it sends ACTIVATION message to its neighboring nodes and continue until the range is covered. All sensor nodes in the network join the activation phase, no matter to which cluster they belong. This eliminates the redundant activation of sensor nodes on the borders of the clusters, which may happen when the activation of nodes is done in each cluster independently.
4.4.2.6 Data communication phase

Once the clusters are formed and active CH sensors are selected, the data communication for localization process begins and the cluster members get localized using HMM.

4.4.3 Cluster Location Estimation using HMM

The HMM method locates the randomly scattered cluster members in the outdoor environment with the help of CH nodes. The area of grid size \( n \times n \) for node follows the movement models to move from one grid to another. The CH estimates the location of cluster members. Ibrahim & Youssef (2011) have estimated HMM parameters such that each state represents a location in the discrete physical observation and an observation from a state represents RSS reading from associated cluster member. During the operational stage, RSS interpretation from each cluster member and the HMM parameters are the necessary input to estimate the most probable sequence of states that results in the estimated location. The cluster member location is obtained using HMM as discussed in Chapter 3 section 3.4.3.

4.5 SIMULATION ENVIRONMENT

In order to evaluate and validate the performance of the HMM, simulation is carried out using NS-2 (2012). A network area of 1000 m × 1000 m is considered. It is divided into small grid of 100 units. The cluster range \( R_{\text{cluster}} = 2 \times R_{\text{sense}} \). Simulation parameters are shown in Table 4.1. Huang & Zaruba (2007a) have considered that 10% of the total nodes are assumed as anchor nodes (CH) and the network area is deployed with 100 nodes. Initially the sensor node is varied from 20 nodes and assumes three anchor nodes for localization. The transmission range is set to 150 m that leads to coverage of 100%. The transition sequence length parameter ‘N’ is
fixed at 100 and the observation sequence length parameter ‘M’ is fixed at 119.

Table 4.1 Simulation Parameters for Cluster HMM

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Area</td>
<td>1000 m x 1000 m</td>
</tr>
<tr>
<td>Antenna Type</td>
<td>Omni Directional</td>
</tr>
<tr>
<td>Propagation Model</td>
<td>Two-Ray Ground</td>
</tr>
<tr>
<td>Speed</td>
<td>2 – 10 m/s</td>
</tr>
<tr>
<td>Packet Size</td>
<td>512 bytes</td>
</tr>
<tr>
<td>Pause Time</td>
<td>5 s</td>
</tr>
<tr>
<td>Traffic type</td>
<td>CBR</td>
</tr>
<tr>
<td>Initial energy</td>
<td>5.1 Joules</td>
</tr>
<tr>
<td>$R_{\text{cluster}}$</td>
<td>50 m</td>
</tr>
<tr>
<td>$R_{\text{sense}}$</td>
<td>25 m</td>
</tr>
<tr>
<td>Clustering Protocol</td>
<td>CPCP, LEACH</td>
</tr>
<tr>
<td>Model</td>
<td>RWM,RWP</td>
</tr>
<tr>
<td>Method</td>
<td>HMM</td>
</tr>
</tbody>
</table>

4.6 PERFORMANCE METRICS

The performance metrics (Wang & Zhu 2009) are defined below.

- Estimation Error: It is the average distance between the estimated location $x_{\text{est}}$ and the actual location $x_i$ of all sensor nodes. The location error is scaled as the percentage of transmission range $r$.

$$\text{Estimation Error} = \left( \frac{1}{N} \sum_{i=1}^{N} \| x_{\text{est}} - x_i \| \right) / r$$  (4.1)
- Control Overhead: It is the total number of control packets transmitted by the anchors to localize an unknown node in each localization process. Assume to localize a node \( n \), \( B_n \) beacons should be transmitted by \( A_n \) anchors. Then the average control overhead for an unknown node is

\[
Control \ Overhead = \frac{\sum_{n=1}^{N} (B_n/A_n)}{N}
\]  

- Average Energy Dissipation: It is the average amount of energy spent by a sensor node during communication in the network.

4.7 SIMULATION RESULTS AND ANALYSIS

The simulation results for Cluster Localization using HMM are analysed to study the effect of node variation, varying transmission ranges for connectivity and various speeds.

4.7.1 Effect of varying the number of nodes

The evaluation on estimation error or localization accuracy over number of nodes with the speed of 10 m/s are analysed for the CPCP-HMM and is compared with LEACH-HMM. Increase in the number of nodes improves the localization accuracy for different movement models as shown in Figures 4.2 and 4.3.

Increase in the node density lowers the estimation error. The error estimate of RWP CPCP-HMM proves to be better because the node pauses for few seconds while the sensor activates fastly to preserve the coverage area. This specifies that based on the function of CPCP and the localization using HMM tends to converge faster when compared with LEACH-HMM.
Figure 4.2  Impact of Node density on Estimation Error for RWM CPCP-HMM

Figure 4.3  Impact of Node density on Estimation Error for RWP CPCP-HMM
The performance of control overhead over node density is shown in Figures 4.4 and 4.5.

**Figure 4.4  Impact of Node density on Control Overhead for RWM CPCP-HMM**

The CH transmits a packet within its range to gather information from the neighbouring node that increases control overhead. Cluster members overhearing this packet reply their known information to CH. The simulation endorses that RWM CPCP-HMM takes less overhead compared to RWP CPCP-HMM because of the random movement of nodes. Overall, the movement model behaves as per the functionality of the model for both LEACH and CPCP.
Figure 4.5 Impact of Node density on Control Overhead for RWP CPCP-HMM

The effectiveness of average energy dissipated with respect to total number of nodes is shown in Figures 4.6 and 4.7. The average energy spent when in movement is more, but actual energy spent in localization process is less. This shows that energy consumption varies due to increase in the node density. It is observed that for different movement model, average energy dissipation gradually decreases for larger density of nodes. The existing LEACH protocol consumes more energy when compared to CPCP because the CH is selected based on the activation time of the sensor rather than waiting for the time slot allotted in LEACH. RWM CPCP-HMM dissipates less energy because this movement model follows the random direction.
Figure 4.6  Impact of Node density on Average Energy Dissipation for RWM CPCP-HMM

Figure 4.7  Impact of Node density on Average Energy Dissipation for RWP CPCP-HMM
4.7.2 Effect of the coverage

The impact of estimation error against coverage to a transmission range of 150 m for 100 nodes with the speed of 10 m/s is shown in Figures 4.8 and 4.9.

Figure 4.8 Impact of Coverage on Estimation Error for RWM CPCP-HMM

Figure 4.9 Impact of Coverage on Estimation Error for RWP CPCP-HMM
The effect of coverage becomes low when the network is dense, i.e., increase in transmission range. The estimation error increases due to increase in the transmission range for higher node density. The network connectivity is efficiently controlled by varying the transmission range. The CH eventually propagate throughout the network of varying transmission ranges and allow cluster members to localize themselves using the state sequence estimation provided by HMM. The estimation error for RWP CPCP-HMM appears to be better when compared with RWM for the coverage area.

The control overhead packets differ, as shown in Figures 4.10 and 4.11, with increase in transmission range for varying movement models.

![Figure 4.10 Impact of Coverage on Control Overhead for RWM CPCP-HMM](image)

The control overhead packet gradually decreases for increasing transmission range for higher nodes. As anticipated, the estimated locations become more accurate as more information is exchanged among neighbors. Both the movement models take more or less the same overhead packets for the coverage area and as usual CPCP-HMM are better when compared to LEACH-HMM.
The impact of coverage over the average energy dissipation shows minor difference as shown in Figures 4.12 and 4.13.

**Figure 4.11** Impact of Coverage on Control Overhead for RWP CPCP-HMM

**Figure 4.12** Impact of Coverage on Average Energy Dissipation for RWM CPCP-HMM
As the beacon node percentage varies over the deployment area, the average energy dissipated indicates that more nodes are localized for varying transmission range. The energy spent in RWM CPCP-HMM localization is less and state sequence obtained through HMM compared to LEACH-HMM.

![Figure 4.13 Impact of Coverage on Average Energy Dissipation for RWP CPCP-HMM](image)

**Figure 4.13 Impact of Coverage on Average Energy Dissipation for RWP CPCP-HMM**

### 4.7.3 Effect of varying speed

The effect of varying speed over the estimation error for 100 nodes is shown in Figures 4.14 and 4.15. All the nodes in the network have communication range of 150 m. As the speed increases, the localization error progressively decreases. The estimation error obtained for RWM CPCP-HMM is lesser than the LEACH-HMM with various moving speeds. RWP pauses for few seconds and chooses the speed to move to the next destination. The length of the state sequence obtained in HMM is considered for varying speed and the cluster localization also matters.
Figure 4.14 Impact of Speed on Estimation Error for RWM CPCP-HMM

Figure 4.15 Impact of Speed on Estimation Error for RWP CPCP-HMM
Variation of speed over the control overhead is shown in Figures 4.16 and 4.17.

**Figure 4.16** Impact of Speed on Control Overhead for RWM CPCP-HMM

**Figure 4.17** Impact of Speed on Control Overhead for RWP CPCP-HMM
The increase in the speed gradually decreases the control overhead. RWM CPCP-HMM has lower overhead when compared to RWP CPCP-HMM. RWP follows the speed and direction based on the pause time. Though RWP pauses for few seconds and chooses the destination, the increase in the maximum speed of the nodes considers yielding more overhead when compared with RWM where the movement is random.

The performance of average energy dissipation for varying speed is shown in Figures 4.18 and 4.19. The energy is gradually reduced due to the increase in speed for RWM. In both the movement models, the energy drops down at higher speed. For RWP, since it pauses for a few seconds to take decision for the next movement to reach the destination, it spends more energy than RWM.

![Graph showing impact of speed on average energy dissipation for RWM CPCP-HMM](image)

Figure 4.18  Impact of Speed on Average Energy Dissipation for RWM CPCP-HMM
Figure 4.19 Impact of Speed on Average Energy Dissipation for RWP CPCP-HMM

Summarizing the above observations of the CPCP-HMM, the estimation error reduces and tends to converge faster for varying node density, various transmission ranges and varying speed for different movement models. These observations show that the state sequence estimates for location of cluster members are more accurate and tends to converge faster by minimizing the power dissipation using various movement models.

4.8 SUMMARY

The CPCP-HMM helps to obtain better location accuracy in the outdoor environment through RSS measurement by two-ray propagation model. The HMM approach exploits the RSS measurements to estimate the position of a cluster member. The network connectivity is controlled by varying the transmission range; hence the traffic is avoided by the state sequence estimation; longer the sequence, better the location accuracy. In
addition, through a comparative simulation study of various movement
models it is observed that RWP CPCP-HMM improves the location accuracy
whereas RWM CPCP-HMM minimizes the energy for varying nodes. The
advantage of the CPCP-HMM is the sensor activation time of CPCP and tends
to converge rapidly by the state sequence provided by HMM, helps to reduce
the estimation error and subsequently consumes low power.

To summarize, the CPCP-HMM is better for cluster localization
rather than LEACH-HMM. The advantage of preferring CPCP is the sensor
nodes cover the entire network area by the cluster range and the sensor
activation to preserve the area. After the CH election, the cluster members
with RSS adapt to localize HMM and tend to converge faster for cluster
localization.

In general, HMM has the limitation of having lengthier state
sequence for localization process. This is overcome by maximizing the model
likelihood probability of HMM parameter to a fitness or objective function. A
new hybrid BFA_PSO algorithm proposed for optimizing the fitness function
to obtain faster and more accurate location estimation of sensor nodes is dealt
in chapter 5.