Abstract—Wireless sensor networks (WSN) consist of a large amount of sensor nodes distributed in a certain region. Due to the limited battery power of a sensor node, lots of energy-efficient schemes have been studied. Clustering is primarily used for energy efficiency purpose. However, clustering in WSNs faces several unattained issues, such as ensuring connectivity and scheduling inter-cluster transmissions. In this paper, we propose a location-based spiral clustering (LBSC) algorithm for improving connectivity and avoiding inter-cluster collisions. It also provides reliable location aware routing paths from all cluster heads to a sink node during cluster formation. Proposed algorithm can simultaneously make clusters in four spiral directions from the center of sensor field by using the location information. Three logical addresses are used for categorizing the clusters into four global groups and scheduling the intra- and inter-cluster transmission time for each cluster. We evaluated the performance with simulations and compared it with LEACH algorithm.

Keywords—Network lifetime, Wireless sensor network, Energy efficiency, Clustering, Inter-cluster collision, Connectivity

I. INTRODUCTION

Recent advances in technology have witnessed an increasing in using wireless sensor networks (WSNs) in many applications, including environmental monitoring and military field surveillance [1]. Unlike general purpose data communication networks, WSNs are typically designed for a specific domain of applications. In these applications, hundreds to thousands of low cost sensors are deployed and periodically report physical information such as temperature, pressure, humidity, light, and chemical activities. Reports from sensors are collected by observers (called sinks). Many WSN applications require only the aggregated value at a certain region. In this case, sensors in different positions in a certain region can collaborate to aggregate their data and more efficiently report their information. Moreover, data aggregation reduces the communication overhead in the network, leads to meaningful energy savings.

In order to support such data aggregation or network topology control, nodes can be partitioned into a number of small groups called clusters. Clustering has been considered as an effective approach for organizing the network into a connected hierarchy. Besides achieving energy efficiency, a well designed clustering mechanism can reduce packet collisions between nodes so that it results in better network throughput under high load conditions. Many localization algorithms [2][3] and clustering approaches [4-9] for WSNs have been proposed.

Each cluster has a coordinator, referred to as a cluster head (CH), and a number of member nodes (MNs). The MNs report sensing data to the respective CHs. The CHs not only perform sensing the environments, but also collect the data from MNs and relay the aggregated data to a sink through gateways (GWs) [4] or other CHs [6][7].

Clustering in WSNs causes several issues, such as ensuring connectivity and scheduling inter-cluster communications [10][11]. In this paper, we propose a location-based spiral clustering (LBSC) algorithm to handle the problems. We iteratively select CHs located along with the four spiral directions. In this paper, we assume that a node knows the location information of its one-hop neighbor nodes (NNs) by exchanging hello messages.

The contributions of LBSC are as follows. First, while generating new clusters by LBSC algorithm, the node sets of four different directions can execute LBSC clustering simultaneously without any control packet collisions. Therefore we can reduce the network-wide clustering latency. Second, after completing the cluster formation, each node will be assigned in one of the four global groups and all clusters assigned to the same global group can concurrently transmit data to their respective CHs without inter-cluster collision. Third, each CH is able to select a GW node among one-hop NNs to configure the routing path to the sink node in the process of cluster formation.

The rest of this paper is organized as follows. In Section II, the motivation and approach of our research is introduced. Section III presents our proposed LBSC algorithm in detail. In Section IV, the simulation results are presented. Finally, Section V concludes this paper.

II. MOTIVATION AND APPROACH

Fig. 1 shows general problems in clustering method. First, after cluster formation, if some of nodes are not contained in any cluster because of the limited transmission range, they need a rejoin method for network connectivity. It is unsuitable for energy efficient design.
Second, a CH can set the time schedule and inform such schedule to its MNs for intra-cluster transmissions. The problem is how to prevent the inter-cluster collisions \((b_2 \rightarrow \text{CH}_B)\) and how to support the concurrent transmission \((a_1 \rightarrow \text{CH}_A)\) as shown in Fig. 1. One possible way is to use different code division multiple access (CDMA) code [6]. However, a relatively expensive cost is required for the usage of CDMA.

One possible way is to use different code division multiple access (CDMA) code [6]. However, a relatively expensive cost is required for the usage of CDMA.

For solving the above-mentioned problems, we propose a LBSC algorithm for improving connectivity and avoiding inter-cluster collisions. The basic idea is to iteratively construct clusters in four spiral directions from the center of the deployed sensor field using the location information as shown in Fig. 2. \(A_1\) denotes direction and \(A_2\) is for the order of clustering tier. Rotating the role of CHs among nodes is essential so as not to lay a burden on a few nodes with more duties than others. The first CH \((\text{CH}_{\text{FST}})\) can be replaced by its NNs in every round and we can select different next tier cluster head \((\text{CH}_{\text{NEXT}})\) by using different degree \((\theta)\).

As shown in Fig. 1, in order to distinguish four global groups, LBSC algorithm uses three logical addresses as follows. The first address, denoted by \(A_1\), is initially set as 0 for the \(\text{CH}_{\text{FST}}\) and increases as a new cluster is generated in each spiral direction, in which \(A_1\) represents clustering tier number from the center as shown in Fig. 2. The second address, denoted by \(A_2\), distinguishes the four different spiral directions. The combination of the \(A_1\) and \(A_2\) distinguishes the four global groups for avoiding the inter-cluster packet collisions. The third address, denoted by \(A_3\), is used for indexing each MN in a cluster. This address can be changed at every round by clustering procedure.

In order to balance the energy consumption between the CHs (including \(\text{CH}_{\text{FST}}\)), the cluster range of the CH which is closer to the center should be smaller than the others. Thus, our proposed algorithm adaptively calculates the proper cluster range \((\text{CR})\) of each CH as follows:

\[
\text{CR}_j = \min\{R, \alpha \times R + \beta \times A_i \times R\}
\]

(1)

where \(R\) represents the predefined maximum transmission range. \(\alpha\) and \(\beta\) are scaling parameters. As a CH is departing from the center, its cluster range is increasing up to the maximum value.

In order to select proper CH, DN, and GW, we use \(\text{NN}_i\) as the set of neighbor nodes (NNs) of the node \(k\) and \(\text{MN}_i\) as the set of MNs of the node \(k\). If a node is located within the \(\text{CR}\) from \(\text{CH}_{\text{FST}}\), it becomes a MN of \(\text{CH}_{\text{FST}}\). Note that \(\text{MN}_i \subseteq \text{NN}_i\). In the following, the node that is nearest to a basis position \(BP\) among the node set \(S\) is derived as (2).

\[
j^* = \arg \min_{j \in S} \sqrt{(BP_x - j_x)^2 + (BP_y - j_y)^2}
\]

(2)

where \(j^*\) represents a node contained in set \(S\); \(\{j_x, j_y\}\) represents the position of node \(j\).

LBSC also establishes routing paths from all CHs to the sink during clustering. Each CH selects a GW node with the location information of the \(\text{NN}_i\), \(\text{CH}_{\text{CUR}}\), and the sink. Each CH can select a GW node nearest to the sink using (2) where \(S\) is \(\text{NN}_i\) and \(BP\) is the sink node position.
III. LOCATION BASED SPIRAL CLUSTERING (LBSC)

LBSC algorithm has the following network conditions. All nodes are fixed and have the same transmission range R. Each node knows the location information of its one hop NNs by using the localization algorithm. In order to prolong the network lifetime, rotating the role of CHs is basically handled by a predefined timer.

A. Time Scheduling

As shown in Fig. 4, in every round, the proposed clustering method is performed at first, and then data transmission and sleep mode are followed repeatedly.

![Figure 4. Super frame structure.](image)

As shown in Fig. 5 and Fig. 6, after node deployment, the closest node from the center of the sensor field announced by the sink node becomes the CHFST. Then it selects the four DNs to determine the CHFST in four different directions. And each CH selects a DN simultaneously. This step will be performed recursively until all nodes are included in clusters.

![Figure 5. Location based spiral clustering algorithm.](image)

![Figure 6. Location based spiral clustering algorithm.](image)

After cluster formation, we can distinguish four global groups and establish time schedule for concurrent data transmissions based on cluster addresses denoted by $A_1,A_2=odd.odd$, $A_1,A_2=odd.even$, $A_1,A_2=even.odd$, and $A_1,A_2=even.even$ as shown in Fig. 6. Then clusters of the same group can simultaneously transmit the data to the respective CHs without inter-cluster collisions. In Fig. 6 the shadowed clusters represent “$A_1,A_2=odd.even$” group and they can use the same time interval for data transmission.

For sensing data transmission to the sink node, each node sends its data to the respective CH. Then the CH aggregates the data and sends it to the sink through GWs. Data transmission with inter-cluster coordination can be implemented as two different manners as shown in Fig. 7. First, the cluster time schedule is divided into four different time period for four global groups. Second, each global group uses different frequency channel for simultaneous data transmission. For both cases, the time period for single cluster is also partitioned to the mini slot assigned to each MN with address $A_1$ category.

![Figure 7. Data transmission with proposed address.](image)

B. Step 1: Selection of the First Cluster Head

The CHFST should be located around the center of the sensor field for successful cluster formation. At the beginning of the proposed clustering algorithm, the sink determines the center position of the sensor field, and then broadcasts a control message, which is carrying the center position and a predetermined range of the possible CHFST from the center. If a node is inside the predetermined range of the center and it does not have any one hop neighbor that is closer to the center, then it will perform the role as the CHFST after random back-off time. CHFST can be replaced by its neighbors in every round for fair energy consumption.

C. Step 2: Selection of the First Tier Decision Nodes

The first work of CHFST is to determine four DNs located in four different directions. CHFST defines four coordinates, $P_i (i = 0, 1, 2, 3)$, as the BP of each direction as follows:

$$P_i = CHFST_i + R \cos(90^\circ \cdot i + \theta)$$

$$P_i = CHFST_i + R \sin(90^\circ \cdot i + \theta)$$

where $C_{center}$ is the number of NNs within predetermined range from center of the sensor field. We can select different CHFST by using $\theta$. In this paper, we use $\omega = 25^\circ$. Therefore, at each round, cluster selection base point is rotating by $\theta$.

CHFST can select the four DNs nearest to the respective $P_i$ among its NNs as in (2) where S is $NN_{CHFST}$ and BP is $P_i$. $A_2$ address of each selected DN follows the index $i$ of each base position $P_i$ as shown in Fig. 8.
D. Step 3: Selection of the Second Tier Cluster Heads

As shown in Fig. 9, in order to select the second tier cluster heads, CHSND, in four directions, each DN defines new coordinate, \( P_i \) \((i=0,1,2,3)\), as the basis position for selecting second clusters.

\[
P_i = CH_{FST} + \frac{4}{3} CR_i \cos(90^\circ \cdot i + \theta)
\]

(4)

The reason of calculating \( P_i \) is to avoid a hole which includes certain nodes that are not contained in any cluster. Then DN can select the CHSND nearest to the respective \( P_i \) among the set \( S = NN_{DN} - MN_{CH_{FST}} \) as in (2). In order to avoid control packet collision, DNs with \( A_2 = (0 \text{ or } 2) \) select CHSND at first and then DNs with \( A_2 = (1 \text{ or } 3) \) select CHSND next as shown in Fig. 9.

E. Step 4: Selection of Decision Nodes for \( A_1 > 0 \)

After selecting four CHSNDs, each CHSND node concurrently repeats selecting DNs and CHs until no more nodes will be found in four directions. Method for selecting the DN is divided into two procedures as follow.

1) Calculating Intersection Points: As shown in Fig. 10, each CHCUR can find two intersection points of the line that crosses the line between CHCUR and CHFST at a right angle and the transmission range circle from CHCUR. Each CH can calculate two intersection points \( \{(x_1, y_1), (x_2, y_2)\}\) by using a quadratic formula. Then select one point \( P \) from two intersection points as in (5). The reason of calculating \( P \) is to select proper DN.

\[
\begin{align*}
\text{if } (CH_{CUR} - x > CH_{FST} - y) \\
\quad \text{if } (y_1 - slope(x_1 - CH_{FST} - x) - CH_{FST} - y > 0) \\
\quad \quad \quad P_x = x_1, P_y = y_1 \\
\quad \quad \quad \text{else } P_x = x_1, P_y = y_2 \\
\quad \text{else } (y_1 - slope(x_1 - CH_{FST} - x) - CH_{FST} - y < 0) \\
\quad \quad \quad P_x = x_1, P_y = y_1 \\
\quad \quad \quad \text{else } P_x = x_1, P_y = y_2
\end{align*}
\]

(5)

where slope is \((CH_{FST} - x - CH_{CUR} - x) / (CH_{FST} - y - CH_{CUR} - y)\).

2) Selecting Decision Nodes: As shown in Fig. 10, in order to select the DN, first CHCUR constructs the SDN (set of DN candidate nodes), in which to be a DN candidate node, any node \( N \) of \( NN_{CH_{CUR}} \) should satisfy the following two conditions as in (6).

\[
\sqrt{(CH_{CUR} - x - N - x)^2 + (CH_{CUR} - y - N - y)^2} > \frac{R}{2}
\]

(6)

Then CHCUR can select the DN nearest to the respective P among the set as in (2) where \( S = S_{DN} \) and \( BP \) is \( P \). If selected DN does not find CHNXT appropriately, it sends a message to the CHCUR to select another DN except it.

F. Step 5: Selection of Cluster Heads for \( A_1 > 1 \)

As shown in Fig. 11, the DN selects the CHNXT with the location information of the NNs, CHCUR, and CHFST. In order to select the CHNXT, DN constructs the set \( S_{CH} = NN_{DN} - MN_{CH_{CUR}} \). Then DN can select the CHNXT, which is the nearest node to the CHFST among the nodes of \( S_{CH} \) using (2), in which \( S = S_{CH} \) and \( BP \) is \( CH_{FST} \).
IV. SIMULATION RESULTS

In this section, we compare the performance of LBSC and LEACH throughout computer simulation. The specific parameters are summarized Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field size</td>
<td>900m × 900m</td>
</tr>
<tr>
<td>Node type</td>
<td>Static node</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>200</td>
</tr>
<tr>
<td>Transmission range of node ((R))</td>
<td>150m</td>
</tr>
<tr>
<td>(\alpha), (\beta)</td>
<td>0.8, 0.05</td>
</tr>
<tr>
<td>(\omega)</td>
<td>25°</td>
</tr>
</tbody>
</table>

Fig. 12 shows that every node in the sensor network is included in one of the clusters generated by LBSC algorithm. We can also see that all the clusters are successfully categorized into one of four global groups. Moreover, we can see that the LBSC algorithm is well formed in spiral directions by following the four colored arrows in Fig. 12. However, in case of LEACH, we can observe that a number of nodes were not able to join any cluster after clustering procedure as shown in Fig. 13.

The result of successful cluster formation using LBSC, each set of clusters are engaged in one of the four global groups as shown in Fig. 14. Thus each set of clusters engaged in a global group can concurrently perform the packet transmission without inter-cluster packet collision using the inter-cluster time schedule.

In Fig. 15, we simulated relatively long time (100 rounds) to evaluate performance. In case of LEACH, we vary the expected number of clusters from 10 to 50. Note that in LEACH, the expected number of clusters should be set before sensor node deployment. Unlike LEACH, in LBSC appropriate numbers of clusters are automatically generated in accordance with node’s transmission range. In Fig. 15-(a), the generated number of clusters in LBSC is about 30 for every round. However in LEACH, even though the expected number of clusters was set to 30, the actually generated clusters are about 40 because LEACH algorithm requires the cluster re-forming procedure for the nodes that fail to participate in any cluster as shown in Fig. 15-(b). It represents that LEACH requires additional packet exchanges, and thus may have the lengthy network-wide clustering latency. Fig. 15-(c) shows the required inter-cluster time slot separations (or the required frequency channels) to avoid inter cluster collisions. LBSC always needs four inter-cluster time slots whereas LEACH needs much more time slots. Moreover, in LEACH to avoid inter cluster collisions, neighboring CHs need to negotiate their time slots. LEACH requires complex computation work to guarantee that in the entire sensor network there is no inter cluster collision. LBSC generally has the stable number of nodes in a cluster than that of LEACH as shown in Fig. 15-(d).

![Figure 12. After cluster formation using LBSC.](image1)

![Figure 13. After cluster formation using LEACH.](image2)

![Figure 14. Concurrently perform the packet transmission without inter-cluster packet collision.](image3)
V. CONCLUSIONS

In this paper, we proposed a location-based spiral clustering algorithm which assures that all nodes are included in clusters without control packet collisions during the cluster formation period. It also provides the routing path from CHs to the sink. We categorize all generated clusters into the four global groups and establish time schedule by using three logical addresses. Clusters belonging to the same global groups can share the same time schedule so that we can increase network throughput and avoid data packet collisions. Simulation results show that each global group concurrently transmits data from MN to respective CH without inter-cluster collisions using four inter-cluster time schedule.

ACKNOWLEDGMENT

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute for Information Technology Advancement) (IITA-2009-C1090-0902-0019)

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. 2009-0079228)

REFERENCES


