Practical Data Transmission in Cluster-Based Sensor Networks

Dae-Young Kim, Jinsung Cho and Byeong-Soo Jeong
Department of Computer Engineering, Kyung Hee University
Yongin 446-701, South Korea
[e-mail: {kimdy, chojs}@khu.ac.kr]
*Corresponding author: Jinsung Cho

Received March 16, 2010; revised April 27, 2010; accepted June 7, 2010; published June 30, 2010

Abstract

Data routing in wireless sensor networks must be energy-efficient because tiny sensor nodes have limited power. A cluster-based hierarchical routing is known to be more efficient than a flat routing because only cluster-heads communicate with a sink node. Existing hierarchical routings, however, assume unrealistically large radio transmission ranges for sensor nodes so they cannot be employed in real environments. In this paper, by considering the practical transmission ranges of the sensor nodes, we propose a clustering and routing method for hierarchical sensor networks: First, we provide the optimal ratio of cluster-heads for the clustering. Second, we propose a *d*-hop clustering scheme. It expands the range of clusters to *d*-hops calculated by the ratio of cluster-heads. Third, we present an intra-cluster routing in which sensor nodes reach their cluster-heads within *d*-hops. Finally, an inter-clustering routing is presented to route data from cluster-heads to a sink node using multiple hops because cluster-heads cannot communicate with a sink node directly. The efficiency of the proposed clustering and routing method is validated through extensive simulations.

Keywords: Wireless sensor network, cluster, energy-efficient, practical radio ransmission range

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2009-0083992) and by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2009-(C1090-0902-0002)).

DOI: 10.3837/tiis.2010.06.002

1. Introduction

Recent advances in wireless communications and electronics have enabled the development of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate in short distances [1]. Sensor networks are composed of a large number of sensor nodes that are densely deployed in a physical space, which is called a sensor field. They monitor physical phenomena, deliver information, and cooperate with neighbor nodes [2]. **Fig. 1** depicts a typical example of wireless sensor networks. Each sensor node has the capabilities to collect data and route data to the sink node. Data are routed to the sink node, which communicates with the Internet, by multi-hop transmission in an infrastructureless area [1]. These sensor nodes should have ad-hoc networking ability, which does not necessitate network infrastructure, to communicate with other nodes. Ad-hoc network schemes cannot be applied directly to sensor networks because sensor networks consist of several nodes and transmit data using broadcast and data-centric features. There have been numerous studies on efficient routing algorithms in wireless sensor networks. A data-centric feature with attribute-based addressing in sensor networks differs from other wireless networks that use IP addresses. The sink node floods queries that specify the features of required data, then sensor nodes respond if the collected data in the sensor nodes corresponds with the queries.

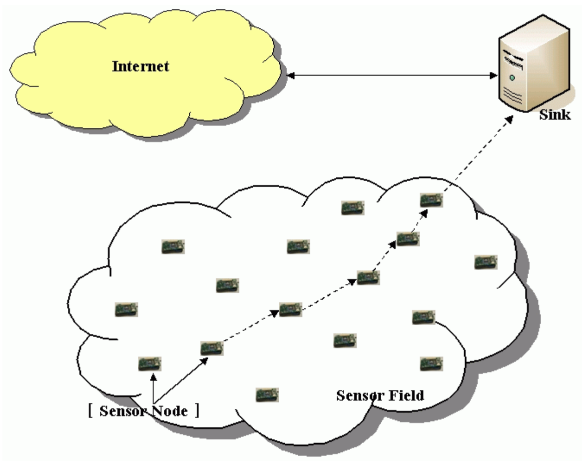
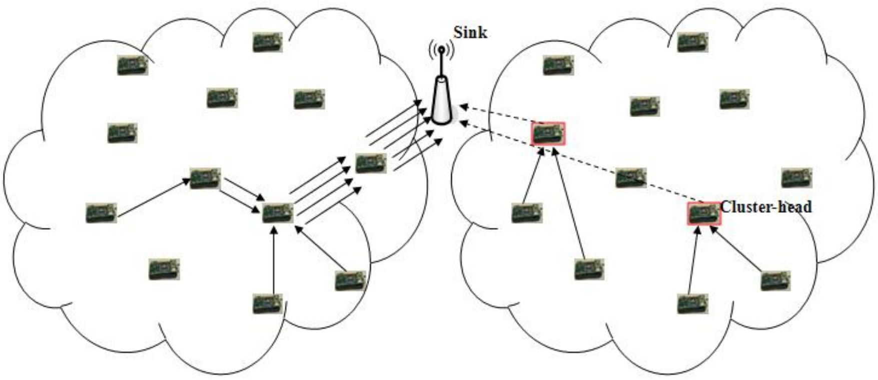


Fig. 1. Wireless sensor networks

In general, routing protocols for wireless sensor networks are classified into two types: flat routing protocols and hierarchical routing protocols. Flat routing protocols are the same as typical data-centric protocols. Data are requested through queries and the properties of the data are specified by attribute-based addressing. The nodes that receive the query have the same opportunity to transmit data and they route data to the sink node through a multi-hop network. Flat routing protocols as data-centric routing protocols follow the aggregation paradigm, whereby data aggregation is performed at intermediate nodes to reduce the number of data transmissions. Addressing schemes such as attribute-value pairs, however, might not be sufficient for complex queries and are usually dependent on applications [3]. In addition, all intermediate nodes must decide how long to wait for data from each of their neighbors. Waiting a long time at intermediate nodes results in more data and thus higher accuracy but transmission delay will be increased [4]. We cannot apply the aggregation paradigm to flat routing protocols without determining an optimal waiting time. Since conventional flat routing protocols do not include algorithms for waiting time, excessive traffic is transmitted over flat routing protocols. That is, although flat routing protocols are practical in large scale sensor networks, a large quantity of data is transmitted in flat routing protocols.

In hierarchical routing protocols, sensor nodes in a sensor field construct clusters for routing and then data transmission occurs as two steps, i.e., intracluster routing and inter-cluster routing. Typically, the energy consumption for data communication is the highest portion of total energy consumption in sensor networks. Because sensor nodes have limited power, they need fewer data transmissions for their long lifetimes. Hierarchical routing protocols have fewer data transmissions than flat routing protocols because the number of whole data transmissions is smaller in the network. Each cluster-head compresses and aggregates data from slave nodes within its cluster. Cluster-based hierarchical routing protocols therefore show better performance than flat routing protocols [4]. The clustering method in hierarchical routing protocols has been widely used in ad-hoc networks and sensor networks, and there are various algorithms to construct a cluster [5][6][7][8][9]. **Fig. 2** shows data transmissions of flat routing and hierarchical routing protocols.



(a) Flat routing (b) Hierarchical routing

Fig. 2. Wireless sensor network routing protocols

Clustering methods are mainly considered in hierarchical routing protocols that have been studied in wireless networks. Conventional hierarchical routing protocols in sensor networks do not consider inter-cluster communication because they assume that cluster-heads can communicate with the sink node directly. However, IEEE 802.15.4 (LR-WPAN) which is one of the transmission standards for wireless sensor networks, is focused on 'Personal Operating Space' (POS[1]) that typically extends up to 10m in all directions [10][11].

Neither cluster-heads nor sensor nodes can transmit data directly over POS. Although there are several studies on heterogeneous hierarchical sensor networks composed of low power sensor nodes and more powerful cluster-heads [12][13], it is unrealistic to suggest that the positions of cluster-heads should be determined a priori. We consider POS for the construction of sensor networks because conventional hierarchical routing protocols cannot be employed in a practical environment.

Fig. 3 depicts the aforementioned problems which assume an unrealistic environment. First,

[1] POS means a space around a person or object.

since transmission radio distance is less than 10m in POS and thus the diameter of each cluster is less than 10m in conventional clustering algorithms, many clusters must be organized. Numerous clusters lead to large communication overhead so that the network lifetime may be reduced. Next, every cluster-head cannot transmit data directly to the sink node because its transmission radius is the same as that of other sensor nodes within the POS.

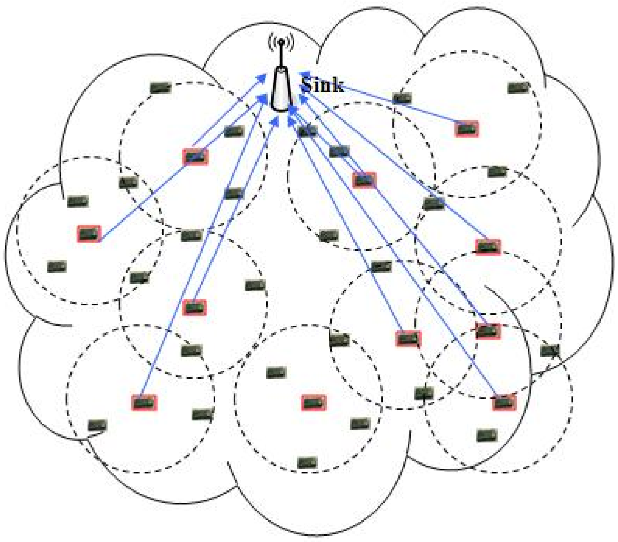


Fig. 3. Problems of hierarchical routing protocls

In this paper, we consider multi-hop transmission for a hierarchical routing in practical environments. First, since the number of cluster-heads affects the performance of sensor networks, we provide a model to obtain the optimal ratio of cluster-heads. Second, we propose a clustering scheme using a d-hop approach to maintain the ratio in a POS environment. Finally, we present an intra/inter-cluster routing based on hop counts in multi-hop transmission. We evaluate the efficiency of the proposed method through extensive simulations based on a realistic POS environment.

The remainder of this paper is organized as follows. We briefly introduce hierarchical routing and flat routing algorithms in wireless sensor networks as related work in Section 2. Section 3 proposes our clustering and intra/inter-cluster routing. Section 4 presents the performance evaluation and we conclude the paper in Section 5.

2. Related Work

Although cluster-based hierarchical routing algorithms can reduce communication costs compared with flat routing algorithms, communication is still a major power consumption factor in hierarchical routing algorithms. There have been many attempts to reduce communication costs by constructing clusters efficiently.

In the most classical method of constructing clusters, identifications of nodes are used. Each node is identified by a unique integer ID (identification) after deployment. The linked cluster algorithm by Baker selects the highest numbered node as a cluster-head in the mobile ad-hoc network [5]. If there are several nodes outside the region of the highest numbered node N, node N-1 becomes a cluster-head. Since the linked cluster algorithm always keeps the highest ID among nodes, each node must know the IDs of other nodes or a centralized node should maintain IDs of all nodes. This method is not appropriate for sensor networks because sensor networks are composed of many sensor nodes and clustering should be done in a distributed fashion. In addition, it may not be efficient because sensor networks exploit random IDs for clustering.

When there are a large number of nodes in sensor networks, in general, we can obtain better performance if we utilize the connectivity of each node. Gerla proposed a clustering algorithm using the connectivity of nodes in a mobile ad-hoc network environment [6]. In this algorithm, the node with the highest connectivity becomes a cluster-head within a cluster, and if nodes have the same degree of connectivity, the node that has the lowest ID becomes a cluster-head. Because a connectivity-based clustering algorithm must maintain the connectivity values of all nodes, it is not the correct method to use in wireless sensor networks where a large number of sensor nodes are deployed. Furthermore, the connectivity of nodes does not change for a long time because sensor nodes do not move. Thus cluster-heads continuously maintain their roles so that high power consumption occurs at specific nodes which are cluster-heads.

Basagni proposed the Distributed Clustering Algorithm (DCA) which is an advanced method for clustering in a mobile ad-hoc network environment [7]. Every node indicates weight as a measure of its importance and it exchanges messages to get the IDs and weights of its neighbors. Then, nodes that have larger weights perform the roles of cluster-heads. The main advantage of this approach is that it is possible to choose more suitable cluster-heads through a distributed method by considering the importance of nodes as weights. However, DCA does not suggest algorithms to assign weights to nodes.

Because all nodes do not exchange their information with all other nodes in wireless sensor networks, clusters should be constructed with minimum information collected from neighbor nodes. In such wireless sensor networks, the representative hierarchical routing algorithm is ‘low-energy adaptive clustering hierarchy’ (LEACH) [8] which selects cluster-heads based on a probabilistic method so that energy consumption can be uniformly distributed among nodes. Each member node of a cluster is in single hop distance apart from its cluster-head and cluster-heads directly communicate with the sink node. In the probability-based clustering methods, however, we cannot guarantee that cluster-heads are uniformly distributed in a network field. Although the fraction of nodes that may become cluster-heads is given, it may be impossible to cover all areas of sensor networks with that fraction when we apply POS considered in LR-WPAN as explained in **Fig. 3**. In addition, the assumption that every cluster-head directly communicates with the sink node is unrealistic because sensor nodes with restricted resources cannot propagate via radio over 10m in a POS environment.

Hybrid energy-efficient distributed clustering (HEED) improves LEACH by considering the residual energy of each node [9]. That is, it selects the node that has the highest residual energy as a cluster-head. The cluster-head node computes the Average Minimum Reachability Power (AMRP) and delivers this AMRP to its neighbor nodes. Neighbor nodes select cluster-heads with lower AMRP. HEED assumes the same energy consumption model as LEACH and tries to maintain the ratio of cluster-heads by changing the transmission radius of nodes. HEED, however, inherits the aforementioned problems of LEACH because it uses the same environment as LEACH. Maintaining the ratio of cluster-heads for clustering through a variable transmission radius is also impossible in a practical POS environment.

While existing hierarchical routing protocols have several constraints in reality because cluster-heads cannot directly transmit data to the faraway sink node, flat routing protocols can be used in real environments since they route data through a multi-hop network. The ‘directed diffusion’ (DD) [14] protocol is the representative flat routing protocol using data-centric characteristics. In DD, the sink node requests data using a query message to sensor nodes then sensor nodes respond if they have the corresponding data. DD has numerous data transmissions in densely deployed sensor networks over a large area. Generally, a sensor node has a correlation with its neighbor nodes regarding sensed data in the local area. Thus there are many nodes with similar data in a dense sensor network and each node with data of interest participates to deliver the data to the sink node. This is the typical weak point of flat routing protocols.

Table 1 summarizes the characteristics of the above protocols presented in this section. As we mentioned earlier, compared with flat routing protocols, hierarchical routing protocols show better performance with regard to energy savings. However, in a real environment with IEEE 802.15.4 POS, there are several problems with hierarchical routing protocols as described in **Fig. 3**, because the radio propagation distance of a sensor node is very short and a sensor node has limited energy. First, sensor networks must have more clusters in order to cover all areas of the network field due to the short radio range of a sensor node in POS. Second, cluster-heads cannot transmit data directly over POS.

Table 1. Clustering and routing protocols in wireless sensor networks

	Classification	Clustering	Cluster-head selection	Transmission
D. J. Baker [5]	Hierarchical	Centralized	Identification	Direct
M. Gerla [6]	Hierarchical	Centralized	Connectivity	Direct
S. Basagni [7]	Hierarchical	Distributed	Weight	Direct
LEACH [8]	Hierarchical	Distributed	Probability	Direct
HEED [9]	Hierarchical	Distributed	Residual Energy	Direct
DD [14]	Flat	Distributed	-	Multi-hop
Proposed	Hierarchical	Distributed	Connectivity + Residual Energy	Multi-hop

In the hierarchical sensor networks, to construct clusters in a distributed manner, the ratio of cluster-heads should be given. Bandyopadhyay et al. [15] and Chen et al. [16] analyzed the ratio but the algorithms from both studies have inaccuracies in the analysis procedure and results. Computation of the ratio by Chen et al. resulted in a high ratio of cluster-heads because they do not consider multi-hop communication and POS. Using the high ratio, their scheme constructs many clusters. Bandyopadhyay et al. calculated the ratio of cluster-heads using a Poisson point process on a Voronoi tessellation. They modeled total energy cost in sensor networks using multi-hop communication and obtained the ratio of cluster-heads required to minimize the cost. In their calculation of the energy costs of transmitting aggregated data in cluster-heads to the sink node, the authors employed the average distance from all sensor nodes to the sink node. As a result, an inaccurate ratio of cluster-heads was calculated. Furthermore, since Bandyopadhyay et al. derived the range of clusters from the inaccurate ratio of cluster-heads, more energy was expended to manage the clusters.

In the next section, we explain a more accurate computation method for the ratio of cluster-heads by improving the method of Bandyopadhyay et al. and describe a d-hop clustering using the ratio to expand the range of clusters. In addition, we provide multi-hop routing to be used in a POS environment.

3. Practical data transmission in hierarchical sensor networks

As explained in Section 1, in general, a hierarchical routing algorithm consists of two phases: clustering and routing. In the proposed method, sensor nodes select cluster-heads and construct clusters in the clustering phase. The routing phase to transmit data can be divided into two steps: intra-cluster routing and inter-cluster routing. Sensor nodes within a cluster deliver data to their cluster-head in the intra-cluster routing step and cluster-heads route data to the sink node in the inter-cluster routing step using the interest message received from the sink node. The clustering phase is repeated periodically in order to balance loads of cluster-heads. After constructing clusters, data transmissions occur in the routing phase.

We propose our hierarchical routing protocol in this section. First, we deal with the analysis of the ratio of cluster-heads for more accurate results in Section 3.1. Next, we describe a clustering method using a *d*-hop approach in Section 3.2. Amis et al. [17] explained how the *d*-hop approach could improve network performance but they did not indicate how to determine the *d* value. We propose a method to find the proper *d* value in our clustering method. Finally, we present the hop-count based multi-hop routing for intra/inter-cluster routing which considers minimum hop-count and residual energy in Section 3.3.

3.1 The ratio of cluster-heads

Both LEACH [8] and HEED [9] suggest that 5% is the ratio of cluster-heads required to build clusters. LEACH obtains this ratio through experimental results and HEED uses the ratio of LEACH. However this ratio is available on the condition of direct communication between cluster-heads and the sink node. To construct clusters efficiently, we need to estimate the ratio of cluster-heads taking multi-hop communication into consideration.

To compute the ratio, we make the following assumptions for a Voronoi tessellation.

- Nodes are distributed according to a homogeneous spatial Poisson process with intensity λ .
- The number of total nodes in a circular area is a Poisson random variable, N with mean λA , where A is area.
- Nodes have probability p of becoming a clusterhead.
- A transmission radius of each node is R_t.
- A wireless channel is a free space and error-free.
- Processing energy in a node is not considered.
- A sink node exists in the sensor field.

The network field is divided into several zones called Voronoi cells. Each Voronoi cell consists of a cluster-head and member nodes of the cluster-head. Cluster-heads are denoted as P1 which has the intensity $\lambda_1 = p\lambda$ and non cluster-heads are denoted as P0 which has the intensity $\lambda_0 = (1-p)\lambda$. The intensities are homogeneous spatial Poisson process.

Using Foss [18][19], we obtain the number of P0 particles (N_c) and the total length of all the segments (L_c) connecting the particles of the P0 to the nucleus P1 in a Voroni cell when the number of nodes n is given

$$E[N_c \mid N = n] = \frac{\lambda_0}{\lambda_1} \tag{1}$$

$$E[L_c \mid N = n] = \frac{\lambda_0}{2\lambda_1^{3/2}}. \tag{2}$$

These equations are derived by aggregate characteristics (S_f). In a Voronoi cell, an aggregate indicates the phenomenaon that P0 particles connect to the nucleus P1 [18][19].

When we consider a circular cell, the S_f function, which is applied to the Campbell theorem[2] [20] and Palm distribution [3] [21], can be represented as

$$E[S_f \mid N = n] = \lambda_0 \int_0^{\infty} f(l) 2\pi l \exp(-\lambda_1 \pi l^2) dl \ , \tag{3}$$

Taking $f(l) = 1$ and $f(l) = l$ we get the expectations of the variable N_c and L_c respectively, where l means the length of each particle. When we define $f(l)$ to describe the hop-count of each particle, we can derive the total hop-count in a Voronoi cell.

The hop-count for each particle depends on l and distances between relay nodes. The distances between relay nodes are determined by positions of relay nodes which are placed in a range of the transmission radius (R_t). For example, in **Fig. 4**, the distance r_i (i=1,2, ..., 5) for each hop has a value between 0 and R_t and the P0 particle has 5hops.

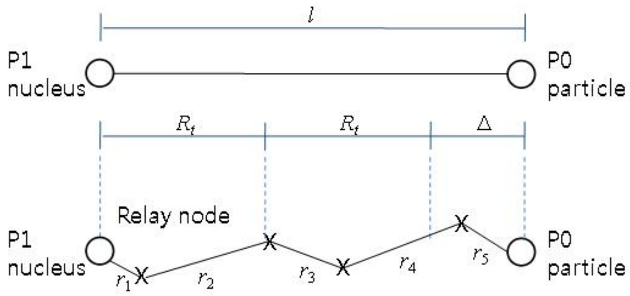


Fig. 4. Relationship between hop-count and positions of relay nodes

Thus we utilize the mean value of the distance betwen relay nodes (r). Using l and r of a particle, the hop-count of a single particle is depicted as $\lceil l/r \rceil$ and the total hop-count (H_c) in a Voronoi cell is

$$E[H_c \mid N = n] = \lambda_0 \int_0^{\infty} \left\lceil \frac{l}{r} \right\rceil 2\pi l \exp(-\lambda_1 \pi l^2) dl \ . \tag{4}$$

$\lceil l/r \rceil$ can be represented by $\dfrac{l}{r} + \alpha$, where α is a value to represent the hop-count with integer $(0 \le \alpha < 1)$. Then Eq. (4) can be written as

$$\begin{aligned} E[H_c \mid N = n] &= \lambda_0 \int_0^{\infty} \left(\frac{l}{r} + \alpha \right) 2\pi l \exp(-\lambda_1 \pi l^2) dl \\ &= \frac{\lambda_0 + 2r\alpha\lambda_0\sqrt{\lambda_1}}{2r\lambda_1^{3/2}} . \end{aligned} \tag{5}$$

We now calculate the total energy consumption cost using Eq. (1) and Eq. (5). Then we derive the probability p required to minimize the total energy cost. The total energy consumption in hierarchical sensor networks occurs in two hierarchies on a Voronoi tessellation. In the first hierarchy, we deal with the relation between a cluster-headand its member nodes in a Voronoi cell. In multi-hop transmission, the energy cost is represented by multiplying 1-hop transmission cost and hop-count. When C_{1st} is the energy cost used by the member nodes to transmit data to the cluster-head, the energy cost is

---

[2] The Campbell theorem represents the number of nodes by density and deployed area of nodes.

[3] Palm distribution for a node x connecting to the cluster-head T(0) with a radius l: $P\{x \in \mathrm{T}(0)\} = \exp(-\lambda_1 \pi l^2)$.

$$
\begin{aligned}
E[C_{1st} \mid N=n] &= E[T_{1-hop}] \cdot E[H_c \mid N=n] \\
&= T_{\cos t} \cdot \frac{(1-p)+2r\alpha(1-p)\sqrt{p\lambda}}{2rp^{3/2}\sqrt{\lambda}},
\end{aligned} \tag{6}
$$

where T_{cost} is the mean value of 1-hop transmission cost $E[T_{1\text{-}hop}]$. Since $E[H_c|N=n]$ indicates total hop-count for data transmission of each node in a Voronoi cell, the energy cost in a Voronoi cell is represented by Eq. (6).

In the second hierarchy, we consider a sensor field, which consists of several cluster-heads and the sink node, as a large Voronoi cell. Given the number of nodes is n, major factors p' and λ' are

$$
p' = \frac{1}{1+np} \tag{7}
$$

$$
\lambda' = \frac{(1+np)\lambda}{n}. \tag{8}
$$

Then, we can denote H_{CH} as the total hop-count of all segments connecting the cluster-heads to the sink node when $\lambda'_0 = (1-p')\lambda'$ and $\lambda'_1 = p'\lambda'$

$$
\begin{aligned}
E[H_{CH} \mid N=n] &= \frac{\lambda'_0 + 2r\alpha\lambda'_0\sqrt{\lambda'_1}}{2r\lambda'^{3/2}_1} \\
&= \frac{(1-p')+2r\alpha(1-p')\sqrt{p'\lambda'}}{2rp'^{3/2}\sqrt{\lambda'}}.
\end{aligned} \tag{9}
$$

The energy cost in the second hierarchy is computed in the same manner as the first hierarchy

$$
\begin{aligned}
E[C_{2nd} \mid N=n] &= T_{\cos t} \cdot E[H_{CH} \mid N=n] \\
&= T_{\cos t} \cdot \frac{\lambda'_0 + 2r\alpha\lambda'_0\sqrt{\lambda'_1}}{2r\lambda'^{3/2}_1} \\
&= T_{\cos t} \cdot \frac{n^{3/2}p + 2r\alpha np\sqrt{\lambda}}{2r\sqrt{\lambda}}.
\end{aligned} \tag{10}
$$

The total energy consumption cost is the summation of the costs which occur in the first hierarchy and the second hierarchy

$$
\begin{aligned}
E[C \mid N=n] &= E[C_{2nd} \mid N=n] + E[C_{1st} \mid N=n] \cdot np \\
&= T_{\cos t} \cdot \left(\frac{n^{3/2}p + 2r\alpha np\sqrt{\lambda}}{2r\sqrt{\lambda}} + \frac{(1-p)+2r\alpha(1-p)\sqrt{p\lambda}}{2rp^{3/2}\sqrt{\lambda}} \cdot np \right).
\end{aligned} \tag{11}
$$

Removing the conditioning on N yields:

$$
\begin{aligned}
E[C] &= E[E[C \mid N=n]] \\
&= E[N] \cdot T_{\cos t} \cdot \left(\frac{\sqrt{E[N]}p + 2r\alpha p\sqrt{\lambda}}{2r\sqrt{\lambda}} + \frac{(1-p)+2r\alpha(1-p)\sqrt{p\lambda}}{2r\sqrt{p\lambda}} \right) \\
&= \frac{\lambda A T_{\cos t}}{2r\sqrt{\lambda}} \cdot \left((\sqrt{\lambda A} + 2r\alpha\sqrt{\lambda})p + \frac{(1-p)+2r\alpha(1-p)\sqrt{p\lambda}}{\sqrt{p}} \right).
\end{aligned} \tag{12}
$$

We can derive the optimal p that minimizes the total energy consumption cost $E[C]$ from Eq. (12).

3.2 *d*-hop Clustering

Given the ratio of cluster-heads, clustering should be performed in a distributed manner. This is a major requirement in hierarchical routing protocols. In the proposed clustering algorithm, all nodes do not need to receive the information required for clustering from a particular central node. Rather, each node obtains status information such as the identifier of a node, the strength of the power signal, and the ratio of residual energy from neighboring nodes before constructing clusters. The information obtained by this procedure is also used in cluster-head selection and intra/inter-cluster routing.

```
Scan Neighbor Nodes
Send ADV message to neighbors
if (receive ADV message)
    Make Neighbor Nodes Table (NNT)

Build Cluster
Compute threshold to select a cluster-head
Generate random number R
if (threshold > R)
    Become a cluster-head
    Send CH_STAT message to the members (until d-hop range)
    Receive JOIN messages
else
    Not a cluster-head
    if (receive CH_STAT message)
        Send JOIN message
    else
        Become a cluster-head
        Send CH_STAT message to the members (until d-hop range)
        Receive JOIN messages
```

Fig. 5. The proposed clustering algorithm

NID	PW_{signal}	$E_{residual}$	*HID*	HC_{intra}	*SID*	HC_{inter}

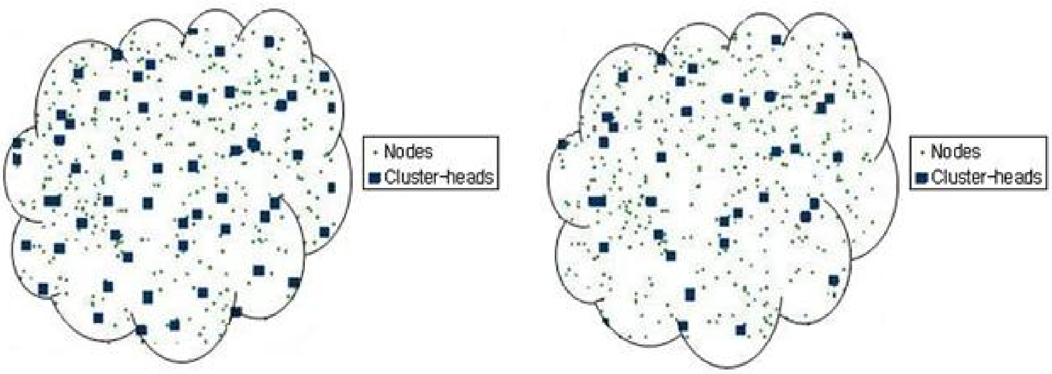
Fig. 6. Neighbor node table: NNT

Fig. 5 describes the pseudo-code of the proposed clustering algorithm. First, the procedure to scan neighbor nodes is performed. Each node sends an ADV message to neighbor nodes through broadcast using the maximum communication range of a single hop in order to advertise its status information; it subsequently makes a Neighbor Nodes Table (NNT). The ADV message includes an identifier and the residual energy of a neighbor node. The strength of the power signal to deliver data to a neighbor node can be indicated by the strength of the power signal of the received ADV message.

In the proposed hierarchical routing, every node contains an NNT as a routing table. In **Fig. 6**, *NID* is an identifier of the sensor node, PW_{signal} is the power strength for transmitting data to a neighbor node, and $E_{residual}$ is the residual energy of a neighbor node *NID*. *HID* is an identifier of a cluster-head, and HC_{intra} is the hop-count from a cluster-head to a current node. *SID* is an identifier of the final destination that denotes the sink node'sidentifier. HC_{inter} is the hop-count from the sink node to a current node. Each sensor node overwrites NNT periodically and the first three fields are filled through scanning neighbor nodes in the clustering phase. The next two fields (*HID*, HC_{intra}) are filled after deciding on appropriate cluster-heads during the clustering phase and are used in intra-cluster routing. The final two fields (*SID*, HC_{inter}) are filled by the *interest* message broadcasted from the sink node and are used in inter-cluster routing.

Next, the procedure to construct clusters is implemented. Each node generates a random number and computes a t*hreshold*. Then, they compare these two values (*threshold* and random number *R*). If the *threshold* is larger than the random number *R*, the node becomes a cluster-head. Both LEACH and HEED use this procedure to select cluster-heads. For the cluster-head selection, the proposed method can employ the existing cluster-head selection scheme. In the proposed algorithm, each selected cluster-head locally broadcasts a CH_STAT message *d* wireless hops away at most. By doing this, the range of a cluster extends *d*-hops from its cluster-head. The CH STAT message includes the identifier of the cluster-head and the hop-count from a cluster-head. Other nodes receiving the CH_STAT message, which are not the cluster-head, fill the information from the message into their NNT'sand send JOIN messages to participate in the cluster as a member node. If a node does not receive the CH_STAT message, the node becomes a cluster-head and broadcasts its CH_STAT message.

In the process of building clusters, the range of clusters influences the network performance. When *d* is 1 in the proposed clustering algorithm, lots of clusters are constructed with conventional clustering algorithms. As mentioned in Section 1, conventional clustering algorithms which construct clusters with a 1-hop range have numerous clusters to cover the entire area in a real transmission environment. When *d* is large, fewer clusters are constructed but the load of cluster-heads required to build and manage clusters increases. Hence, hierarchical routing protocols should keep an optimal number of clusters for efficiency. We also need to determine a proper value of *d* for the proposed clustering algorithm.



(a) The result of 1-hop clustering (b) The result of 3-hop clustering

Fig. 7. Selected cluster-heads (500 sensor nodes)

Fig. 7 shows cluster-heads selected for clustering when 500 sensor nodes are deployed. **Fig. 7(a)** is the result of 1-hop clustering and **Fig. 7(b)** is the result of 3-hop clustering. As shown in **Fig. 7**, by expanding a range of clusters, a network can manage the number of clusters. Although the optimal ratio of cluster-heads is given from Section 3.1, many clusters are constructed in a POS environment. In hierarchical sensor networks, it is very important to have an adequate number of clusters to maintain energy efficiency. If an appropriate number of clusters is used, by extending the range of clusters until *d*-hop, energy consumption for communication in hierarchical sensor networks can be minimized and the network lifetime will increase. Thus, in the proposed clustering algorithm, we find the proper *d*-hop for a range of clusters given the ratio of clusters (p), the transmission radius (R_t), the number of total nodes (N), and the radius of the sensor field (R_v).

Each Voronoi cell has only one P1 nucleus.

$$\pi R_c^2 \lambda_1 = 1, \tag{13}$$

where R_c is a radius of the cluster. From Eq. (13), R_c can be represented as

$$R_c = \frac{1}{\sqrt{\pi\lambda_1}} = \frac{1}{\sqrt{\pi p \lambda}}. \tag{14}$$

As mentioned earlier, the intensity λ is the number of total nodes divided by the area as

$$\lambda = \frac{N}{\pi R_v^2}. \tag{15}$$

Then, R_c can be rewritten as

$$R_c = \frac{R_v}{\sqrt{Np}}. \tag{16}$$

On the other hand, R_c can be represented by multiplying the mean distance of single hop r and the hop-count of the cluster (d) as depicted in **Fig. 4**. Then, the hop-count is

$$d = \left\lceil \frac{R_c}{r} \right\rceil = \left\lceil \frac{R_v}{r\sqrt{Np}} \right\rceil. \tag{17}$$

Since r ranges between 0 and R_t, $E[r]$ is $R_t/2$. When R_v, R_t, N, and p are given, the hop-count is calculated as

$$d = \left\lceil \frac{2R_v}{R_t\sqrt{Np}} \right\rceil. \tag{18}$$

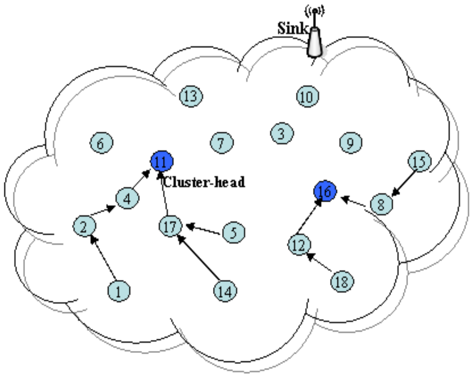
3.3 Hop-count based routing

After selecting cluster-heads and constructing clusters, intra/inter-cluster routings are required to transmit data to the sink node. The routing algorithm is based on NNT. Every sensor node contains a special data structure NNT as a routing table. The status information of neighbor nodes, identifiers of cluster-heads, and hop-counts from cluster-heads are filled during the clustering phase. In addition, information about the identifier of the sink node and hop-count from the sink node are also filled from the received *interest* message which the sink node broadcasts to the sensor field. After completing NNT, each node routes data through its NNT. Since a node which has fewer hop-counts than a current node is always selected in NNT as the next hop, a routing loop does not occur in the proposed algorithm. NNT is updated periodically when clusters are constructed and the $E_{residual}$ field of NNT can be changed during clustering phases or by receiving information about the change of the ratio of residual energy of a neighbor node.

3.3.1 Intra-cluster routing

Sensor nodes use *HID* and HC_{intra} fields in NNT when they communicate with their cluster-head. HC_{intra} contains hop-count information between a cluster-head and its member nodes within the cluster and both *HID* and HC_{intra} are written into NNT when sensor nodes receive the CH_STAT message from other nodes. The CH_STAT messages are routed until they are d hops away from their cluster-heads and HC_{intra} is incremented by one when they are routed. Sensor nodes receive the CH_STAT messages and update *HID* and HC_{intra} of the messages to the NNT. Each sensor node looks up its NNT when it transmits data to the neighbor node having the lowest HC_{intra}. If there are several nodes where the lowest HC_{intra} is the same, the nodes deliver data to the neighbor which has the highest Eresidual. In addition, when sensor nodes transmit data, they use PW_{signal} which is obtained during the clustering phase.

Fig. 8 shows an example of intra-cluster routing. First, node 1 refers to NNT to find node 2 which has the lowest HC_{intra} from itself to a cluster-head. Node 2 performs the same operation and delivers data to node 4. Finally, node 4 transmits data to its cluster-heads.



NNT

	NID	PW_{signal}	$E_{residual}$	HID	HC_{intra}	SID	HC_{inter}
Node1:	2	20	89	11	2	0	6
Node2:	4	24	65	11	1	0	5
	1	20	91	11	3	0	7
Node4:	11	18	71	11	0	0	4
	17	20	78	11	1	0	5
	2	24	89	11	2	0	6

Fig. 8. An example of intra-cluster routing

3.3.2 Inter-cluster routing

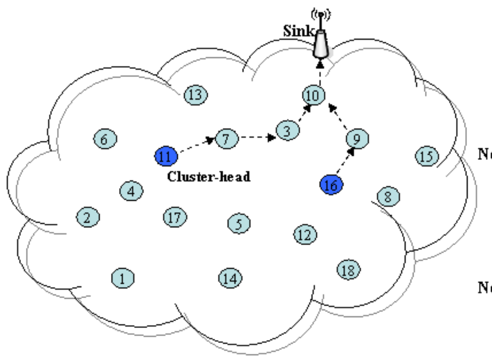
As mentioned earlier, cluster-heads deliver data to the sink node directly in conventional routing protocols. However, this is not possible when we consider the practical transmission radius for sensor nodes. Cluster-heads should route data by multi-hop routing through other sensor nodes. Thus, we use the hop-count based inter-cluster routing algorithm, which is the same as intra-cluster routing except for referenced fields in NNT, to transmit data to the sink node. In inter-cluster routing, HC_{inter}, which is the hop-count information from the sink node to a current node, is needed; we get this information from the *interest* message. When sensor nodes receive the *interest* message, the nodes know the identifier of the sink node and HC_{inter}, which is incremented by one when the *interest* message is delivered from the sink node, and the last two fields of NNT are updated. The proposed inter-cluster routing algorithm is similar to intra-cluster routing except that cluster-heads use SID and HC_{inter} instead of HID and HC_{intra} as routing information.

Fig. 9 is an example of inter-cluster routing. Data from the cluster-head, node 11, are transmitted in the following sequences in the same way: node 11, node 7, node 3, node 10, and the sink node. The final node 10 checks whether SID is NID or not and sends data to the sink node if they are equal to each other.

4.1 Performance metrics

Our proposed hierarchical routing algorithm aims at efficient operation in a real environment. We develop new performance metrics that reflects the practical sensor network environment. Based on a combination of our performance metrics and traditional metrics, we compare the performance of our proposed algorithm with the existing hierarchical algorithms.

Traditional measurements employ 'first node die' (FND) and 'last node die' (LND) metrics [8][9]. FND is the network lifetime to present the degree of load balance about energy consumption. LND indicates the general network lifetime. However, data transmission to the sink node should be done using multi-hop routing in practical wireless sensor networks. Thus, when all the relay nodes near the sink node die, cluster-heads in hierarchical sensor networks cannot deliver data to the sink node. So, it is meaningless to measure LND in a real environment. In this paper, we define another metric 'connection nodes die' (CND) which is the amount of time that neighbor nodes (connection nodes) of the sink node are alive to make connections between cluster-heads and the sink node. If all connection nodes die, the sink node cannot receive sensor data from the network field. We use CND and FND for performance metrics. To represent these metrics in the simulation, *round* is used which is a popular unit for a network lifetime in hierarchical sensor networks. The *round* consists of a set-up phase to organize the clusters and a steady-state phase in which data gathering of the sink node occurs several times [8].



NNT

	NID	PW_{signal}	$E_{residual}$	HID	HC_{intra}	SID	HC_{inter}
Node11:	7	19	81	11	1	0	3
	6	17	79	11	1	0	5
	4	18	65	11	1	0	5
	17	16	78	11	1	0	5
Node7:	3	20	80	11	2	0	2
	11	19	71	11	0	0	4
	13	17	68	11	2	0	4
Node3:	10	24	82	16	2	0	1
	7	20	81	11	2	0	2
Node10:	0	18	-	-	-	0	0
	3	24	80	11	2	0	2
	9	20	78	16	1	0	2

Fig. 9. An example of inter-cluster routing

4. Performance evaluation

4.2 Simulation environments

Table 2. Parameters for performance evaluation

Parameters	Value
Radius of the network field R_v	50 m
Data packet size	100 bytes
Query packet size	25 bytes
Header packet size	25 bytes
E_{elec}	50 nJ/bit
E_{amp}	10 pJ/bit/m^2
E_{init}	1 J
Position of the sink node	Center of the network field
# of data gathering of the sink node in a *round*	5
Transmission radius of a node R_t	10 m

In the previous section, we proposed a practical method for clustering and routing. Now, we evaluate the performance of our method and compare it with those of LEACH [8], HEED [9], Bandyopadhyay et al. [15], and Chen et al. [16] through extensive simulations. The simulator for the evaluation has been implemented in C++. Both LEACH and HEED are representative hierarchical routing algorithms and the algorithms of both Bandyopadhyay et al. and Chen et al. provide the ratio of cluster-heads which should be used in hierarchical routing algorithms. However, since the algorithms cannot be directly compared with the proposed algorithm in a practical environment, we partially modify them.

For cluster-head selection in the clustering phase, we assume that the proposed method, Bandyopadhyay et al., and Chen et al. exploit a residual energy based scheme such as HEED. In addition, because cluster-heads in conventional hierarchical routing algorithms cannot communicate directly with the sink node, we modify them so that they deliver data using our NNT as a routing table. Member nodes of each cluster using existing algorithms transmit data to their cluster-heads directly in intra-cluster routing because the algorithms construct clusters with a single hop range. Cluster-heads route data to the sink node by referencing NNT in inter-cluster routing. We define *LEACH2* as modified LEACH and *HEED2* as modified HEED. Similarly we define *BAND2* as a modified scheme of Bandyopadhyay et al. and *CHEN2* as Chen et al.'s modified scheme. Since Bandyopadhyay et al. and Chen et al. have no cluster-head selection and data routing schemes, we apply our cluster-head selection and data routing to them. In addition, since Bandyopadhyay et al.'s method extends a cluster's range but Chen et al.'s method does not, we apply our *d*-hop computation method to *CHEN2*. For simplicity, we assume an error free wireless environment and randomly distributed sensor nodes in the network field.

In the experiment, we set the number of sensor nodes in a network to 300 and 500. Sensor nodes are deployed in the circular area with 50m radius and the sink node is placed at the center of the network field. Each sensor node has 1 joule as its initial energy and 10m as its maximum transmission distance. We employ LEACH's radio model as an energy consumption model for data transmission. Using LEACH's radio model, to transmit k-bit message for distance s, the radio of sender expends ($E_{elec} * k + E_{amp} * k * s^2$)J and the radio of receiver expends ($E_{elec} * k$)J [8]. Basic environment parameters are the same as HEED [9].

4.3 Simulation results

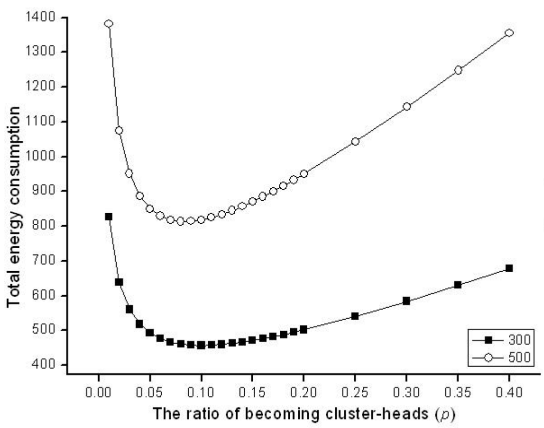


Fig. 10. Total energy consumption following the ratio of becoming cluster-heads

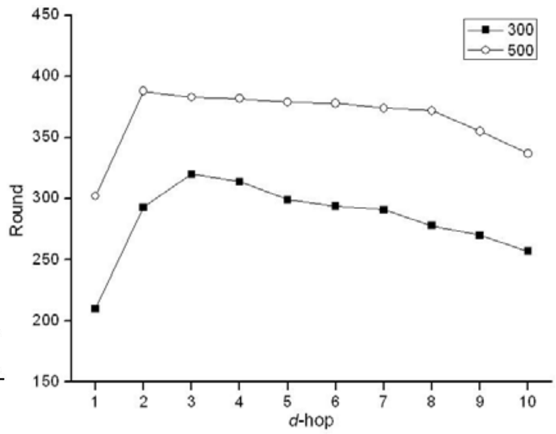


Fig. 11. Network lifetime (FND) with varying d-hop values

First, the ratio p of cluster-heads is estimated to compare algorithms. From Eq. (12), we obtain the ratio p as 0.0837 for 500-nodes and 0.1003 for 300-nodes. From **Fig. 10**, it is clear that these ratios minimize energy expenditure on data communication.

Given the ratio p, the number of nodes N, the transmission radius R_t, and the radius of the sensor field R_v, we can determine the proper *d*-hop for a range of clusters to maintain the proper number of clusters for the ratio *p*. **Fig. 11** presents the performance variation of our hierarchical routing according to varying *d*-hop values. When R_v is 50m and *p* is 10% and 8.37% for 300 and 500 nodes, respectively, Eq. (18) computes *d*-hop values of 3 and 2, respectively. In the simulation, sensor nodes and cluster-heads are randomly distributed. Nevertheless the *d*-hop values in **Fig. 11** are equivalent to Eq. (18). This shows the *d*-hop computation can be applied to the real environment.

Hierarchical routing algorithms employ parameters such as the ratio of cluster-heads (*p*) and *d*-hop for the range of clusters in **Table 3** as major factors for simulation. The factors *p* and *d* of *Proposed* are calculated by Eq. (12) and Eq. (18). *LEACH2* and *HEED2* get their factors from LEACH and HEED. *BAND2* obtains *p* and *d* from Bandyopadhyay et al. However, *CHEN2* gets only *p* from Chen et al. Thus we calculate *d* from *p* in *CHEN2* through Eq. (18).

Table 4 compares the average number of cluster-heads required for hierarchical routing algorithms. In the case of *Proposed* and *BAND2* which expand the range of clusters, fewer clusters are made and they maintain clusters with the ratio *p*. In contrast, for *LEACH2* and *HEED2*, although we set *p* = 5%, many clusters are constructed and the ratio of cluster-heads has values greater than 5%. Since sensor nodes have a short transmission range in a real environment, clusters constructed with a single hop range cannot cover the entire sensor field. To cover the whole area, more clusters are required. In addition, *CHEN2*, which has a high *p* ratio constructs numerous clusters.

Table 3. System parameters for simulation

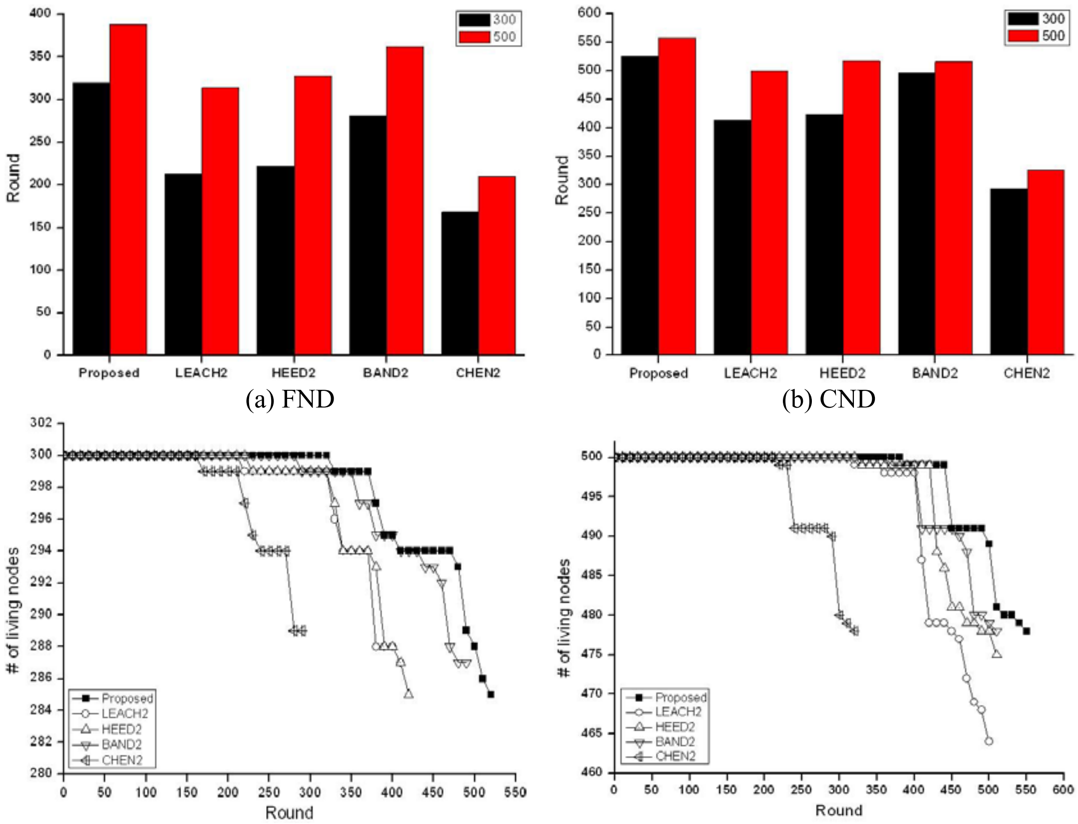
	300 nodes		**500 nodes**	
	p	***d***	***p***	***d***
Proposed	0.1003	3	0.0837	2
LEACH2	0.05	1	0.05	1
HEED2	0.05	1	0.05	1
BAND2	0.1214	5	0.1012	4
CHEN2	0.276	2	0.232	2

Table 4. Average number of cluster-heads on hierarchical routing

	300 nodes	**500 nodes**
Proposed	30.4	43.1
LEACH2	49	53.3
HEED2	48.3	50.4
BAND2	33.7	48.5
CHEN2	76.8	110.7

Fig. 12 illustrates a network lifetime of hierarchical routing algorithms when the number of nodes is 300 and 500. Among the algorithms, *Proposed* and *BAND2* have a longer network lifetime based on both FND and CND metrics because they maintain the number of clusters as the given ratio of cluster-heads for clustering by expanding the range of clusters until *d*-hop. *Proposed*, however, is more efficient than *BAND2* because *Proposed* exploits the more accurate ratio of cluster-heads (*p*) and *d*-hop to organize clusters than BAND2 as mentioned in Section 3. The proposed algorithm not only provides the longest FND and CND but also maximizes the number of alive nodes as shown in **Fig. 12(c)**, **(d)**. This is because *Proposed* maintains the optimal number of clusters.

Table 5 presents collected data at the sink node during CND. It is straightforward that the amount of collected data per *round* is similar in all algorithms because they can deliver all the data during their lifetime. Since CND is the network lifetime during which the sink node can receive data from the sensor field, longer CND indicates the sink node can collect more data. This is why we have presented CND in this paper. As shown in **Fig. 12** and **Table 5**, the amount of collected data increases proportionally to CND.



(a) FND (b) CND

(c) Change in # of alive nodes: N=100 (d) Change in # of alive nodes: N=500

Fig. 12. Performance comparison: Network lifetime

Table 5. Amount of collected data

	During its lifetime		Per *round*	
	300 nodes	**500 nodes**	**300 nodes**	**500 nodes**
Proposed	782989	1388722	1491.4	2488.7
LEACH2	616738	1239341	1493.3	2478.7
HEED2	632723	1283937	1492.3	2483.4
BAND2	740126	1284507	1492.2	2489.3
CHEN2	438118	809072	1495.3	2489.4

5. Conclusions

Sensor nodes have a limited radio transmission radius in a practical sensor network environment. Because existing hierarchical routing algorithms assume that data transmission from cluster-heads to the sink node and from sensor nodes to their cluster-heads can be accomplished in one hop, we cannot adopt these algorithms to real environments. Therefore, in this paper, we have proposed a practical method for clustering and routing in hierarchical sensor networks. The proposed method is efficient because of the following features. It provides the optimal ratio of cluster-heads for clustering. It maintains an appropriate number of clusters using a d-hop approach. In addition, to efficiently route sensor data, it is energy-aware for multi-hop routing. From our analysis, the optimal ratio of cluster-heads and an appropriate d-hop are the dominant factors in the performance improvement observed.

Experimental results validated that the proposed algorithm can improve the network lifetime as much as 27.1% (number of nodes=300), 11.6% (number of nodes=500) for the case of CND compared with *LEACH2*. We can gather as much as 26.9% (number of nodes=300) or 12.1% (number of nodes=500) more data in the sink node compared to *LEACH2* or *HEED2*. In a practical sensor network environment, therefore, the proposed algorithm is an excellent candidate for use in clustering and routing algorithms.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102-114, 2002.

[2] D. Culler, D. Estrin, and M. Srivastava, “Guest editors’ introduction: Overview of sensor networks,” *IEEE Computer*, vol. 37, no. 8, pp. 41-49, Aug. 2004.

[3] K. Akkaya and M. Younis, “A survey on routing protocols for wireless sensor networks,” *Journal of Ad Hoc Networks*, vol. 3, no. 3, pp. 325-349, May 2005.

[4] H. Karl, “Protocols and Architectures for Wireless Sensor Networks,” John Wiley & Sons, 2005.

[5] D. J. Baker and A. Ephremides, “The architectural organization of a mobile radio network via a distributed algorithm,” *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1694-1701, Nov. 1981.

[6] M. Gerla and J. T. Tsai, “Multicluster, mobile, multimedia radio network,” *ACM/Baltzer Journal of Wireless Networks*, vol. 1, no. 3, pp. 255-265, Aug. 1995.

[7] S. Basagni, “Distributed clustering for ad hoc networks,” in *Proc. of IEEE ISPAN 1999*, pp. 310-315, Jun. 1999.

[8] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in *Proc. of 33rd Hawaii Int. Conference on System Sciences*, Jan. 2000.

[9] O. Younis and S. Fahmy, “Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach,” in *Proc. of IEEE INFOCOM 2004*, vol. 1, pp. 629-640, Mar. 2004.

[10] J. A. Gutierrez, M. Naeve, E. Callaway, M. Bourgeois, V. Mitter, and B. Heile, “IEEE 802.15.4: A developing standard for low-power low-cost wireless personal area networks,” *IEEE Network*, vol. 15, no .5, pp .12-19, Sep./Oct. 2001.

[11] E. H. Callaway, “Wireless Sensor Networks Architectures and Protocols,” Auerbach Publications, 2004.

[12] G. Gupta and M. Younis, “Load-balanced clustering of wireless sensor networks,” in *Proc. of IEEE ICC 2003*, vol. 3, pp. 1848-1852, May 2003.

[13] A. Bogdanov, E. Maneva, and S. Riesenfeld, “Power-aware base station positioning for sensor networks,” in *Proc. of IEEE INFOCOM 2004*, vol. 1, pp. 585-596, Mar. 2004.

[14] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed diffusion: a scalable and robust communication paradigm for sensor networks,” in *Proc. of ACM MobiCom 2000*, pp. 56-67, Aug. 2000.

[15] S. Bandyopadhyay and E. J. Coyle, “An energy efficient hierarchical clustering algorithm for wireless sensor networks,” in *Proc. of IEEE INFOCOM 2003*, vol. 3, pp. 1713-1723, Mar. 2003.

[16] Y. P. Chen, A. L. Liestman, and J. Liu, “Energy-efficient data aggregation hierarchy for wireless

sensor networks," in *Proc. of IEEE International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, Aug. 2005.
[17] A. D. Amis, R. Prakash, T. H. P. Vuong, and D. T. Huynh, "Max-min d-cluster formation in wireless ad hoc networks," in *Proc. of IEEE INFOCOM 2000*, vol. 1, pp. 32-41, Mar. 2000.
[18] S. G. Foss and S. A. Zuyev, "On a voronoi aggregative process related to a bivariate poisson process," *Advances in Applied Probability*, vol. 28, no. 4, pp. 965-981, Dec. 1996.
[19] S. G. Foss and S. A. Zuyev, "On a certain segment process with voronoi clustering," *INRIA, Rapport de Recherche*, May 1993.
[20] D. Stoyan, W. S. Kendall, and J. Mecke, "Stochastic geometry and its applications," 2nd Edition, Wiley, 1995.
[21] J. Kingman, "Poisson processes," Oxford University Press, 1993.

Dae-Young Kim received his B.S. degree in electronics engineering, M.S. degree in computer engineering from Kyung Hee University, South Korea in 2004 and 2006, respectively. He is currently a Ph.D. Candidate in computer engineering at Kyung Hee University, South Korea. His research interests include mobile networking & computing, wireless sensor networking, and embedded systems.

Jinsung Cho received his B.S., M.S, and Ph.D. degrees in computer engineering from Seoul National University, South Korea, in 1992, 1994, and 2000, respectively. He was a visiting researcher at IBM T.J. Watson Research Center in 1998 and a research staff at SAMSUNG Electronics in 1999~2003. Currently, he is an assistant professor of Department of Computer Engineering at Kyung Hee University, South Korea. His research interests include mobile networking & computing, embedded systems & software.

Byeong-Soo Jeong received his B.S. degree in computer engineering from Seoul National University, South Korea in 1983, M.S. degree in computer science from the Korea Advanced Institute of Science and Technology, Korea in 1985 and Ph.D. degree in computer science from Georgia Institute of Technology, Atlanta, USA in 1995. In 1996, he joined Kyung Hee University, South Korea. He is now an associate professor at the College of Electronics & Information at Kyung Hee University. From 1985 to 1989, He was a research staff at the Data Communications Corp., Korea. From 2003 to 2004, he was a visiting scholar at Georgia Institute of Technology, Atlanta, USA. His research interests include database systems, data mining, and mobile computing.