무선 센서 네트워크에서의 에너지 효율을 위한 클러스터 헤더 재배치 알고리즘

An Energy-efficient Clustering Algorithm using the Guaranteed Minimum Coverage for ClusterHeads in Wireless Sensor Networks

김남훈[1] 박태림[1] 권욱현[1] 김정준[2], 김용호[2], 신영희[2], [1]서울대학교 전기컴퓨터공학부 [2]KT 컨버전스 연구소
Nam hun Kim[1], Taerim park[1], Wook hyun Kwon[1]
[1]*School of Electrical Engr. and Computer Science, Seoul National University*
[2]*KT* convergence Laboratory

Abstract

In this paper, a new clustering algorithm using the Guaranteed Minimum Coverage (GMC) is proposed. In the new protocol, an appropriate distribution of clusterheads is accomplished by guaranteeing a stochastic coverage at each clusterhead(CH)s. Using this protocol, the communication cost from clusterheads to their member nodes and the load variance in each clusterheads are reduced. Therefore, the network lifetime can be extended and the fair energy consumption for all CHs can be achieved

Keywords

clusterhead, energy consumption, load balance

I. INTRODUCTION

Recently, advances in technology have enabled the development of small, inexpensive and low-power sensor nodes[1]. The most significant challenge in the design of these wireless sensor networks(WSNs) is energy constraint because all nodes are powered by small batteries and operated in unattended. This constraint requires energy efficient techniques something like cluster- ing algorithms. For efficient communication between nodes, sensor networks are typically grouped into clusters, where each cluster has a clusterhead(CH). CHs are responsible for data aggregation in their cluster, and send the collected data to the base station(BS)[2, 3].

In the clustering algorithm, an appropriate distribution of CHs can reduce the average distance between CH and cluster member nodes. If a node can control the transmiss- ion power, properly distributed CHs yield the reduction of the communication energy consumption. Also the fair energy consumption of each CH can be achieved by balancing the number of nodes in each cluster. Therefore, distribution of CHs is important issue in the clustering algorithm.

This paper is focus on of the distribution of CHs and proposes the new clustering algorithm to extend the network lifetime by pursuing the appropriate distribution of CHs. this is accomplished by guaranteeing the minimum coverage range at each CH.

The total communication cost and the load variance metrics are used to measure the performance. The total communication cost means the gross of communication cost between all CHs and their member nodes. And the load variance means the variance of the number of member nodes at each cluster. To obtain more precise data, it is a assumed that all nodes have always data to send its CH, and can control its transmission range. Therefore, the aggre- gation cost is related to the average link distance between CHs and its member nodes, and the load variance is related to the number of member nodes in each cluster.

II. Related Work

Many clustering algorithms in various contexts have also been proposed in the past, but none of these algorithms consider the formation of CHs. Most of these algorithms assume that CHs are distribute properly, or obtain its formation using location information through GPS system.

Clustering Algorithm problem can be divide into two parts. One is the election method of CHs including the problem that how many the number of CH will be proper. The other is the operation method of cluster. Clusterhead must aggregate data from its member nodes, so appropriate aggregation method and the communication between CHs must be considered.

In this paper, only 1-hop clusters are consi- dered. Therefore, following related algorith- ms are generate 1-hop clusters. All of these algorithms require synchronized clocks and have a complexity of O(n).

In the Linked Cluster Algorithm[4], a node becomes the clusterhead if it has the highest identity among all nodes within one hop of itself or among all nodes within one hop of one of its neighbors. This algo- rithm was improved by the LCA2 algorith- m[5], which generates a smaller number of clusters. The LCA2 algorithm elects as a clusterhead the node with the lowest id among all nodes that are neither a cluster- head nor are within 1-hop of the already chosen clusterheads. The algorithm propos- ed in [6], chooses the node with highest degree among its 1.hop neighbors as a clusterhead. In[7], the authors propose a distributed algorithm that is similar to the LCA2 algorithm.

The Weighted Clustering Algorithm (WC- A) elects a node as a clusterhead based on the number of neighbors, transmission power, battery-life and mobility rate of the node The algorithm also restricts the num- ber of nodes in a cluster so that the perfor- mance of the MAC protocol is not degra- ded. The

Distributed Clustering Algorithm (DCA) uses weights associated with nodes
to elect clusterheads [7]. These weights are generic and can be defined based on the application. It elects the node that has the highest weight among its 1-hop neighbors as the clusterhead. The DCA algorithm is suitable for networks in which nodes are static or moving at a very low speed.

In this paper, the operation of cluster is based on LEACH (Low Energy Adaptive Clustering Hierarchy)[1]. LEACH is desig- ned for sensor networks where an end-user wants to remotely monitor the environment. In LEACH protocol, during the set-up phase CHs are stochastically selected and the clusters are organized. In order to select CHs each node determines a random num- ber between 0 and 1. If the number is less than a threshold, the node becomes a CH for a while. And LEACH incorporates
randomized rotation of the high-energy CH position among the nodes to avoid draining the battery of any one node in the network.

However, it is obvious that previous method of CH election will not automatic- ally lead to minimum energy consumption during data aggregation for a given set of nodes. It is possible that all CHs are locat- ed in an edge of the network or adjacent nodes can become CHs. In these cases, some general nodes have to incerase its transmission power to reach a CH.

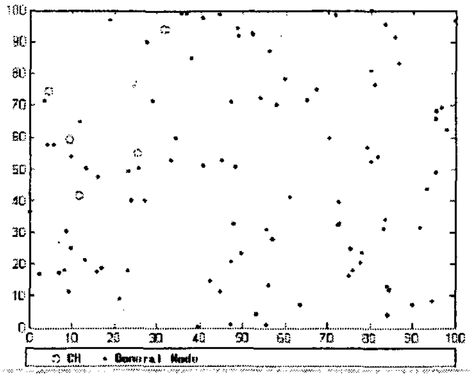

Fig. 1 Bad scenario of CH selection

As an example, consider a case of Figure (1). In this scenario, CHs are selected unfavorably getting together in a left-top corner. Therefore, the total aggregation cost is extremely increased, because the general nodes located in the right part of field must send data to its far-flung CH. Also load variance is increased, because a CH which is surrounded by others has a few member nodes, but the CHs located in outer of CH group must charge on many member nodes.

Increased total communication cost means that the average link distance between CHs and their member nodes is increased. Because all nodes can adjust their transmit power, the increased link distance means that general nodes dissipate more transmis- sion energy to send data to CH.

Load variance has relation to the energy consumption of a CH. If a CH has many member nodes, it consumes more energy to data aggregation and combining. Therefore, the load balance between CHs is needed for the fair energy consumption for CHs.

III. Guaranteed Minimum Cover- age Algorithm

The rotation of responsibility for being CH is needed to achieve the fair energy-consumption as discussed in [1], however, it can not prevent the CHs from getting together in a certain location. In this case, total aggregation cost and load variance are increased. Without the location information of all nodes, it may be impossible to find the optimal formation of CHs. However, energy saving can be achieved by guaranteeing an appropriate coverage of CHs.

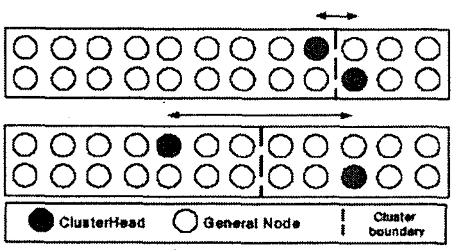

Fig. 2 Appropriate Coverage of CH

As an example, consider cases in Figure 2. The first figure shows that CHs are getting together in right corner. However, if CHs are away as some distance, that is, each CH has a appropriate coverage depicted as in the second figure, the total aggregation cost and load variance can be reduced. Therefore, the worst case as discussed in section 3 can be prevented.

1. Determination of the Guaranteed Minimum Coverage

The Left-parts of Figure 3 shows cases that number of CHs is 4 and 9 respectively. In these cases, the monitoring field can be divided into square areas, so the stochastic coverage radius of CHs can be obtained easily.

However, consider an another case that 5 CHs exist, it is impossible to make a square cluster areas to fit the field. Therefore, generalization through suitable approximation is needed. Equation (1) shows this approximation, and it can be proved straightforwardly as Equation (2).

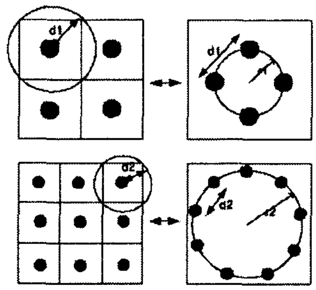

Fig. 3 Coverage and Physical meaning

In general, if there are randomly distributed N nodes in $A \times A$ size field with the probability of being CH λ, there are $N\lambda$ CHs. The first equation in Equation (1) shows the generalize guaranteed minimum coverage(d_{GMC}) obtained from the square area in cases that there exist 4 or 9 CHs. And if $N\lambda$ is large enough (i.e.,over 3 or 4) following approximation can be effected.

$$d_{GMC} = \frac{A}{\sqrt{2N\lambda}} \quad d_{GMC} \propto \frac{1}{\sqrt{N\lambda}} \cong c \cdot \sqrt{n\lambda} \cdot \sin(\frac{\pi}{N\lambda}) \quad (1)$$

where c is a constant.

$$\sqrt{N\lambda} \cdot \sin(\frac{\pi}{N\lambda}) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{N\lambda}} \cdot \left\{ \sin(\frac{\pi}{N\lambda}) / \frac{\pi}{N\lambda} \right\} \Leftrightarrow c \simeq \frac{1}{\pi} \quad (2)$$

Using the Equation (1) and (2),the relation between d_{GMC} and $N\lambda$ can be described as Equation (3):

$$d_{GMC} = \frac{A}{\pi} \cdot \sqrt{\frac{N\lambda}{2}} \cdot \sin(\frac{\pi}{N\lambda}) \quad (3)$$

The right-parts of Figure 3 shows the physical meaning of Equation (3). In Equation (3), d_{GMC} is proportional to $\sqrt{N\lambda}\sin(\pi / N\lambda)$ when the $N\lambda$ is large enough, so using this relation, CHs can be disposed on the circle line maintaining the distance obtained from the left-parts. The radius of circle r is the variable related to the number of CHs and the field size as following equation (4).

$$d_{GMC} \cong 2 \cdot r \cdot \sin(\frac{\pi}{N\lambda}) \Leftrightarrow r = \frac{a}{2\pi} \cdot \sqrt{\frac{N\lambda}{2}} \quad (4)$$

2. ClusterHead Selection

During the network initialization, elected CHs broadcast advertisement(ADV) message after setting the transmission radius as guaranteed minimum coverage (ie, d_{GMC}) in equation (3). In ADV, information of the random value is inserted, and it is used as the priority order. The purpose of this procedure is to find out whether CHs are get together or not. General nodes do not send Join Request yet. In this procedure, general nodes only remember the number of ADVs received from CHs. If a general node hears only one ADV, it becomes the exclusive member of that CH. If CHs are close to each other, these CHs will hear ADVs of others'. In this case, to guarantee the minimum coverage between CHs, except one CH, CHs should be moved out of the ADV range of a remaining CH. Selection of a remaining CH is based on the random value information. If a CH selects lower random value than others', it has a higher priority. Because the mobility of nodes is not considered, CHs needed to move out of the ADV range of a remaining CH, that is a node which has a highest priority, pass the responsibility of being CH to a certain general node. Selection of that general node is limited to their exclusive members. This limitation ensures to keep the guaranteed minimum coverage between other CHs though one of exclusive members becomes CH. Passing the responsibility is done by changing the random value of each other.

Figure 4 shows a algorithm of new CH selection. There are three CHs, called A,B and C. They broadcast ADV in the radius of the guaranteed minimum coverage. In this case, B and C hear ADV of each other. Then they compare the random value in the ADV with the random value of itself, and assume that C has greater value, then C sends GIVEUP message also in the range of the guaranteed minimum coverage. In this figure, C has 4 exclusive members. Once exclusive members hear the GIVEUP message, these nodes send VOLUNTEER message to C. Then C selects one node which has the highest priority among them, if this node is D, then C and D change the random value each other. Once the appropriate CHs is selected, they collect their cluster member nodes. The range of the ADV to guarantee the minimum coverage between CHs may be too short, so there are many orphan nodes which do not receive any ADV.

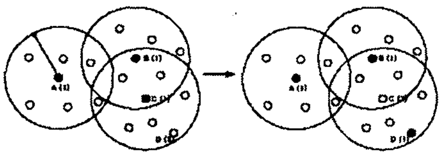

Fig. 4 Procedure of a New CH Selection

Therefore, the re-ADV range to collect member nodes must be increased (ie, up to the maximum transmission range).

3. Selection of the Number of Clusters

The number of clusters can be determined analytically by computing the non-orphan probability of general nodes. In the clustering algorithm, if there exist some general nodes which does not receive the ADV, these nodes can not participate in a cluster. Therefore, These nodes will be isolate (ie, orphaned), and the information from sensing areas of these nodes can not be aggregated. The non-orphan probability means that the fraction of general nodes joined in any cluster. It has intimate relation to the coverage of monitoring field if nodes are densely distributed.

Assume that N nodes are randomly distributed in $A \times A$ sized field with CH probability λ and its transmission range is r .If the boundary condition doesn't considered, the neighboring probability with a reference node is

$$p = \frac{\pi r^2}{A^2} \quad (5)$$

Also the probability that a reference node has k neighbors is described as Equation (6).

$$P_n[n = k] = C_N^k \cdot p^k \cdot (1-p)^{N-k} \quad (6)$$

Therefore, the probability that a reference node does not orphaned can be described as Equation (7).

$$P_n[non-orphan|n=k] = (1-\lambda)[(1-(1-\lambda)^k) \cdot C_N^k \cdot p^k \cdot (1-p)^{N-k}] \quad (7)$$

The first term $(1 - \lambda)$ means that a reference node is not a clusterhead, and $(1-(1-\lambda)^k)$ means that there exist a clusterhead with its neighbors at least. By integrating all cases of k, the non-orphan probability can be obtained as Equation (8).

$$P[Non-Orphan] = \sum_k P_n[non-orphan|n=k] = (1-\lambda)\sum_k (1-(1-\lambda)^k) \cdot C_N^k \cdot p^k \cdot (1-p)^{N-k} \quad (8)$$

If N and A goes to infinity, the probability that a reference node has k neighbors approaches to the poisson distribution as Equation (9).

$$M = E[n] = \frac{\pi \cdot r^2}{A^2} \cdot N = p \cdot N, \quad P_n[n=k] = \frac{M^k \cdot e^{-M}}{k!} \quad (9)$$

Equation (8) must do infinity sum, however, if k is large enough, the probability that a reference node has k neighbors can be approximate as 0. Therefore, let set the Limit Number that can guarantee the non-orphan probability to be high enough (ie, 0.95).

$$P[n \geq k] = 1 - P[n < k] = 1 - \sum_{i=0}^{k-1} \frac{M^i \cdot e^{-M}}{i!} \quad (10)$$

If a k makes the value in equation (10) to be smaller that 0.05, that k is the Limit Number. Therefore, the infinite summation in equation (8), can be reduce to the summation up to Limit Number.

IV. Simulation

The simulation environment is as follows:

- Field size : 100m x 100m
- Number of nodes : 100
- Probability of being CH : 5 %
- Maximum transmission range : 50m

Figure 5 shows the result of the proposed algorithm. Circles are the CHs selected by using the original random value. Asterisks are newly selected CHs adapting the proposed algorithm. In the original case, CHs are getting together in the middle of top area of field, the total aggregation cost and the load variance are extremely increased. However, after adapting the proposed algorithm, these two metrics can be reduced. In figure 5 three CHs are changed, and one of exclusive members of those CHs becomes a new CH.

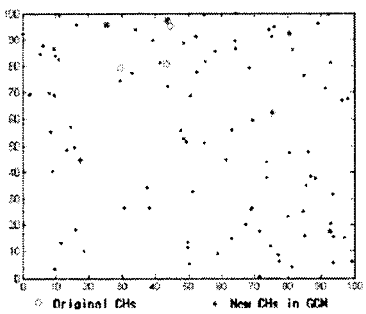

Fig. 5 New ClusterHead Selection

Figure 6 shows the difference of the total link distance. These values mean the saved link distance at each simulation time(round). There exist some rounds that the difference value is below zero. However, at these rounds, the negative values are small and it is limited just a few rounds. The horizon line indicating roughly 200 is the average saved link distance at each round. This is almost 10% of the total link distance in the original protocol.

Total aggregation cost can be obtained from these results. In general, power consumption to transmit a signal over a distance d is proportional to d^n, where $2 \le n < 4$.The exponential n is closer to four for low-lying antenna and near ground channels, as is typical in sensor network communication. Therefore, 10% reduction of average link distance means that almost 20% reduction of transmission power cost at least (ie, set n=2).

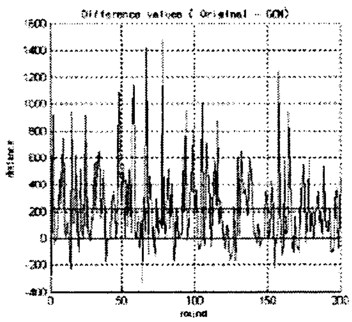

Fig. 6 Saved Total Link Cost

Figure 7 shows the comparison of load variance at equal average condition. The load variance is reduced by using the GMC algorithm. The dotted line is the variance of original one's and the solid line is the variance of GMC's. To make the distribution become uniform as possible, the number of nodes is increase to 600. The load variance of proposed algorithm almost keep a value of 10\% smaller than the original's. Varying the radius of the GMC causes the different total aggregation cost results because the formation of CHs is changed. If the radius is too smaller, the CHs can't be prevented from getting together. Also if the guaranteed is too large, the CH which has the highest priority order takes large area alone. Therefore, the appropriate range of GMC is needed.

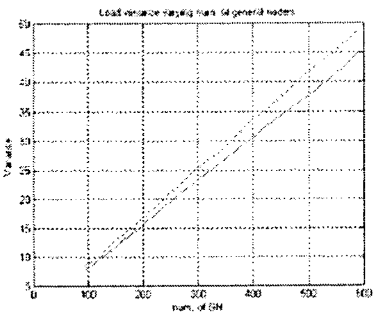

Fig. 7 Comparison of Load Variance

Figure 8 shows the normalized radius of GMC using the Equation (3). Asterisk indicates a value obtained from the square area (i.e., the case that 4,9,16,25 etc. CHs exist). The radius can be found by multiplying the value in Figure 8 and field size of monitoring field. For example, if there exist 5 CHs, the normalized value is nearly 0.3. And if the field size is 100m, then the radius of the GMC is 30m.

Table 1: Accuracy and modification

$N\lambda$	Analysis	Simulation	Compansation
4	32	36	36
5	29.74	34	33.45
6	27.7	32	31.17
7	25.9	29	29.22
8	24.49	27	27.55
9	23.21	26	26.1

Table 1 shows the accuracy of the Equation (3). Values in the column of Simulation show the minimum total link distance at given environment. These values are roughly 10% higher than the analysis. These errors are caused by the random deployment of nodes. For compensating this factor, the constant c (i.e., 1/pi or 0.32) in the Equation (3) becomes 0.36. The third column of Table 1 shows the modification results.

Figure 9 shows that the non-orphan probability. The solid line shows simulation results and the dotted line shows the analysis results. Differences between these two results are caused by the boundary effect. That is, the neighboring probability p in section 4.3 can be slightly smaller if the boundary effect is considered. Simulation shows that non-orphan probability is almost 0.95 when the probability of being CH is 0.05.

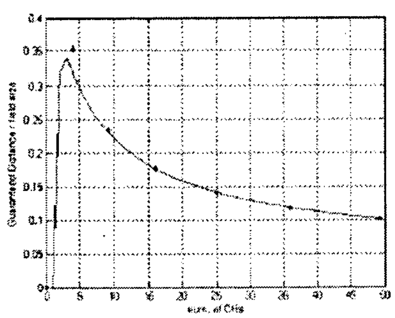

Fig.8 Normalized Guaranteed Minimum Coverage

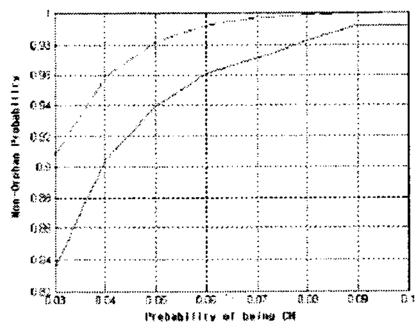

Fig. 9 Non-Orphan Probability

IV. Conclusion

In WSNs, energy-constraint is the most critical issue. This paper proposes the new distribution algorithm of CHs using the guaranteed minimum coverage between CHs. Using this GMC algorithm, the average aggregation cost can be reduced and the load balance can be accomplished. Therefore, the network lifetime is extended and the fair energy consumption for all nodes can be achieved.

Also GMC algorithm can be adaptable to any clustering algorithm which does not use a GPS system. At the initial time to select CHs, proposed algorithm give a chance to achieve a better formation of CHs, and makes the network more energy-efficiently.

For a consideration of the real-time operation and dalay of setup time, GMC algorithm only uses 1 time. However, if the application is not time critical, and setup period is not frequent, the GMC algorithm can be adaptive iteratively. This makes the formation of CHs more better one.

[Reference]

[1] Wendi B. Heinzelman, "An Appli- cation Specific Protocol Architecture for Wireless Microsensor Networks," *PhD Thesis, Massachusetts Institute of Technology*, June, 2000.

[2] Alan D. Amis and Ravi Prakash, "Load-Balancing Clusters in Wireless Ad Hoc Networks," *In Proceedings 3rd IEEE Symposium on Application-Specific Systems and Software Engineering Technology*, pp.25-32, 2000.

[3] Matthias Handy, Marc Haase, and Dirk Timmermann, "Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-Head Selection," *MWCN*, 2002.

[4] 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.

[5] J.E. Wieselthier A. Ephremides and D. J. Baker, "A Design concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling," *Proceeding of IEEE, Vol. 75, No. 1*, pp. 56-73, 1987.

[6] A. K. Parekh, "Selecting Routers in Ad-Hoc Wireless Networks," *Proceedings of ITS*, 1994.

[7] C. R. Lin and M. Gerla, \Adaptive Clustering for Mobile Wireless Networks," *Journal on Selected Areas in ommunication, Vol. 15*, pp. 1265-1275, 1997.