2004년도 한국정보통신설비학회 하계학술대회 논문집 (2004.8.27-28)

무선 센서 네트워크에서의 에너지 효율을 위한 클러스터링 알고리즘

김남훈, 박태림, 권욱현, 장병수*, 김용호*, 이병양*
서울대학교 전기.컴퓨터 공학부 제어정보 시스템 연구실, KT 컨버전스 연구소 통합단말개발팀*

An Energy-efficient Clustering algorithm using the Guaranteed Distance for Wireless Sensor Networks

N.H. Kim, T.R. Park, W.H. Kwon, B.S. Chang*, Y.H. Kim*, B.Y. Lee*
Control Information System Lab. Seoul National University, KT convergence Lab.*

Abstract - In this paper, a new clustering algorithm using the Guaranteed Distance is proposed. In the new algorithm, the appropriate distribution of clusterheads is accomplished by guarantee the stochastic average distance between clusterhead(CH)s.

Using this algorithm, 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 nodes can be achieved.

1. Introduction

Recently, advances in technology have enabled 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 clustering 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 transmission power, properly distributed CHs yield the reduction of the communication energy consumption. Also it can balance the number of nodes in each cluster so fair energy consumption for all nodes can be achieved. Therefore, CHs distribution is important issue in the clustering algorithm. There are many clustering based protocols, but there's no approach to propose the method of CHs distribution.

This paper is focus on the problem of the distribution of CHs and proposes the new clustering algorithm to extend the network lifetime by pursuing the appropriate distribution of CHs. An appropriate distribution of CHs is accomplished by guaranteeing the stochastic average distance between CHs.

The total link cost and the load variance metrics are used to measure the performance. The total link cost means the gross of all link distances between CH and their member nodes. And the load variance means the variance of the number of member nodes at each cluster.

2. Main Discouse

2.1 Problem Formulation

Whether the criterion of CH election is the random value or the number of neighbors, the distribution of CHs is not considered. In the view point of the distribution of CHs, those two methods are similar to the random selection of CHs without any information. It is obvious that the random CH selection will not automatically lead to minimum energy consumption during data transfer for a given set of nodes. It is possible that all CHs are located near the edges of the network or adjacent nodes can become CHs. In this case some nodes have to transmit long distances to reach a CH.

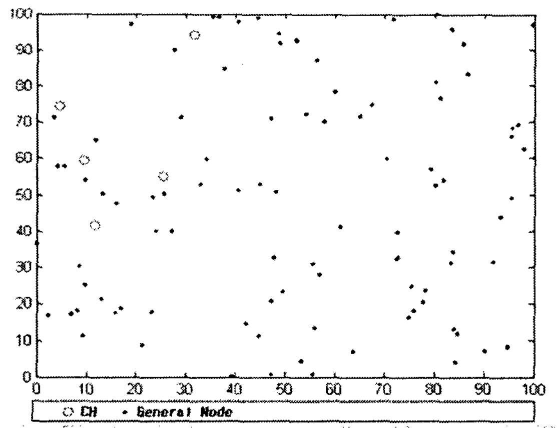

Figure 1. Bad Scenario of CH selection

As an example, consider the case of Figure 1. In this scenario, CHs (cirles) are selected unfavorably getting together in a certain location. In this case, the total link 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 the CH which is surrounded by the others has a few its member nodes, but the CHs located in outer of CH group have many member nodes.

Increased total link 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 transmission energy to send data to CH. Increased load variance metric has relation to the energy consumption of a CH. If a CH has more member nodes, it consumes more energy to data aggregation and combining. Therefore the load balance between CHs is needed for the fair energy consumption of CHs.

2.2 Guarnateed Distance Algorithm

As discussed in section 3, the random CH selection algorithm is needed to fair energy consumption for all nodes, however it can't prevent the CHs from getting together in acertain location. In this case, total link 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, it can be achieved the save of energy by guarantee an appropriate distance between CHs.

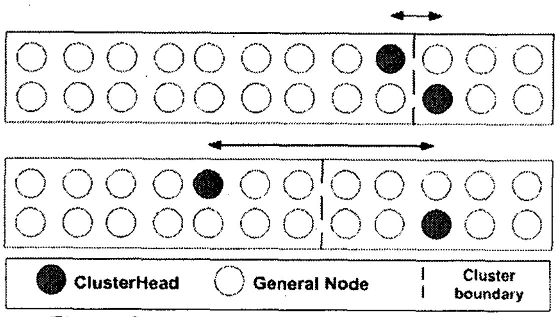

Figure 2. Guaranteed Distance between CHs

As an example, consider the case in Figure 2. The first figure shows the case that CHs are getting together in right corner. However, if CHs are away as some distance depicted as in the second figure, the total link cost and load variance can be reduced. That is, the worst case as discussed in **section 2.1** can be prevented.

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 (i.e., guaranteed distance)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).

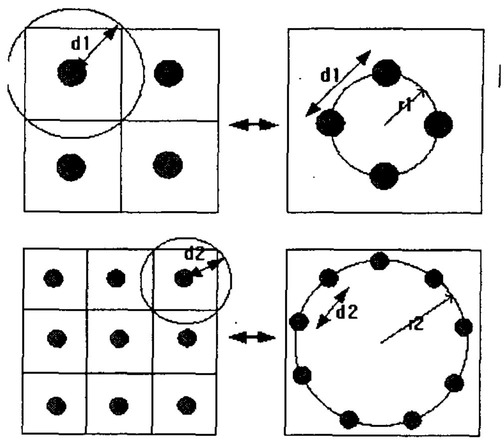

Figure 3. Coverage and formation of CHs

In geleral, if there are randomly distributed 'n' nodes in 'a' by 'a' size field with the probability of selection of CH 'p', there are 'np' CHs. The left equation in Equation (1) shows the generalize guaranteed distance(d_GD) obtained from the square area in cases that there exist 4 or 9 CHs. And if 'np' is large enough (i.e., over 3 or 4) following approximation can be effected.

$$d_{GD} = \frac{a}{\sqrt{2np}},\ d_{GD} \propto \frac{1}{\sqrt{np}} \cong c \cdot \sqrt{np} \cdot \sin(\frac{\pi}{np}) \quad (1)$$

,where c is a constant.

$$\sqrt{np} \cdot \sin(\frac{\pi}{np}) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{np}} \cdot \left\{ \sin(\frac{\pi}{np}) \Big/ \frac{\pi}{np} \right\} \Leftrightarrow c \cong \frac{1}{\pi} \quad (2)$$

Using the Equation (1) and (2),the relation between d_GD and 'np' can be described as Equation (3):

$$d_{GD} = \frac{a}{\pi} \cdot \sqrt{\frac{np}{2}} \cdot \sin(\frac{\pi}{np}) \quad (3)$$

The right-parts of Figure (3) shows the physical meaning of Equation (3). Using the relation in Equation (3), 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).

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

During the network initialization, elected CHs broadcast advertisement(ADV) after setting the transmission radius to the guaranteed distance obtained in Equation (3). In ADV, the random value information 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-REQ yet. In this procedure, general nodes 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 other CHs. In this case, to guarantee the distance 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 the 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 CH to a certain general node. Selection of that general node is limited to their exclusive members. This limitation ensures to keep the guaranteed distance between other CHs though one of exclusive members becomes CH. Passing the responsibility is done by changing the random value of each other.

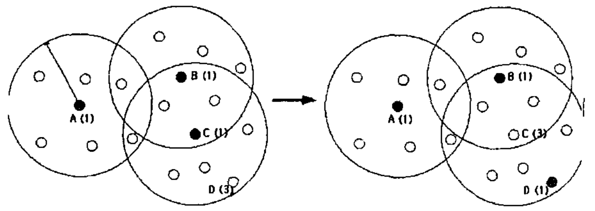

Figure 4. Procedure of a New CH Selection

Figure 4 shows a new CH selection algorithm. There are three CHs, called A,B and C. They broadcast the ADV in the guaranteed distance range. 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 if C has greater value, then C sends GIVEUP message in the guaranteed distance range. 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 the node which has the highest priority, 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 distance between CH may be too short, so there are many orphan nodes which do not receive any ADV. Therefore, the re-ADV range to collect member nodes must be increased.

2.3 Simulation Results

The simulation environment is as follows:

- field size : 100m x 100m
- randomly distributed 100 nodes
- probability of being CH : 5%

Figure 5 shows the result of the proposed algorithm. Circles are the CHs selected by using the 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 link distance 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.

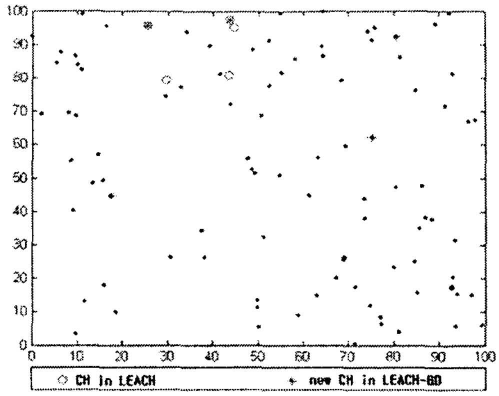

Figure 5. New Clusterheads Selection

Figure 6 shows the difference of the total link distance. This value means the saved link distance at each simulation time(round). There exist 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.

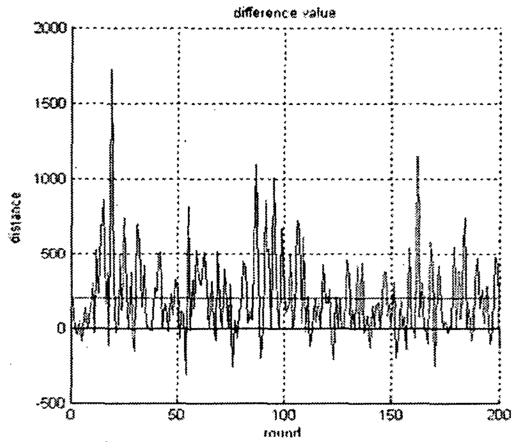

Figure 6. Saved Total link cost

Figure 7 shows the comparison of load variance at equal average condition. The load variance is reduced by using the proposed algorithm. 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 guaranteed distance yields the different total link cost results because the formation of CHs is changed. If the guaranteed distance is too small, 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 guaranteed distance is needed.

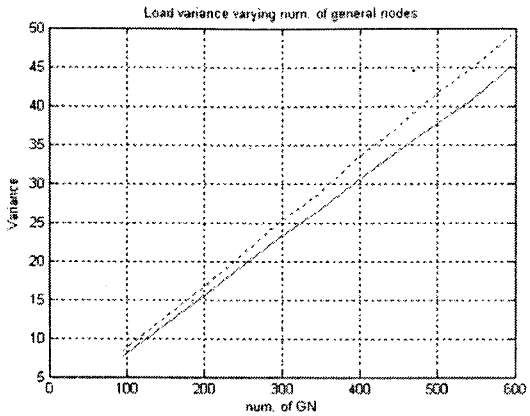

Figure 7. Comparison of Load Variance

Figure 8 shows the normalized guaranteed distance 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 guaranteed distance 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 guaranteed distance is 30m.

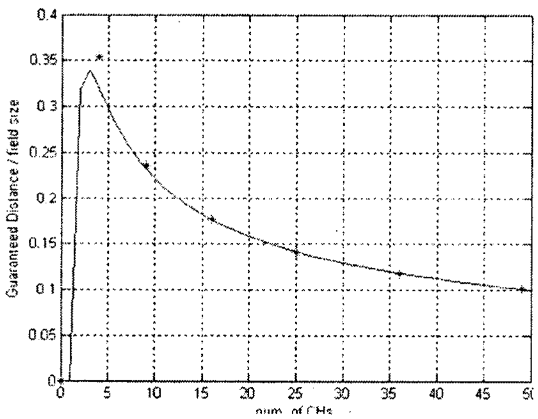

Figure 8. Normalized Guaranteed Distance

Table 1 shows the accuracy of the Equation (3). Values in the column of Simulation result shows the minimum total link distance at given environment. These values are roughly 10% higher than the calculated values. This error is cause by the random deployment of nodes. For compensating this factor, the constant 'c' (i.e., 1/pi or 0.32) in the Equation (3) need to be larger. The modification value of 'c' is 0.36. Table 2 shows the modification result.

3. Conclusion

In WSNs, energy-constraint is the most critical issue. This paper proposes the new distribution algorithm of CHs using the guaranteed distance between CHs. Using this algorithm, the average communication link distance 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. As a future works, it is planned to enhance the algorithm for reduction of load variance.

Table 1. Accuracy of GD

np	Calculated Results	Simulation Results
4	32	36
5	29.74	34
6	27.7	32
7	25.9	29
8	24.49	27
9	23.21	26

Table 2. Modification

np	Calculated Results	Simulation Results
4	36	36
5	33.45	34
6	31.17	32
7	29.22	29
8	27.55	27
9	26.1	26

[REFERENCES]

[1] Wendi B. Heinzelman, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," PhD Thesis, Massachusetts Institude 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 Soft- ware 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.*