2009 World Congress on Computer Science and Information Engineering

Distributed Clustering Algorithm with Load Balancing in Wireless Sensor Network

Farruh Ishmanov, Sung Won Kim
Department of Information and Communication Engineering,
Yeungnam University
Gyeongsangbukdo, Korea

Abstract

Clustering is a promising technique for prolonging network lifetime and improving functionality of a wireless sensor network. It achieves high energy efficiency by utilizing multihop mode communication among clusters. However, relay clusters die much earlier than other cluster, because of unequal burden. As a consequence, network lifetime is significantly decreased. To mitigate this problem, rotation and unequal clustering schemes are proposed. However, rotation schemes are not much energy efficient. Unequal clustering schemes consider only distance and rarely consider overlapping clusters which can seriously affect load balancing. We propose load balancing and energy efficient clustering algorithm to efficiently distribute distance among clusters considering data volume and set up targeted clusters. In each step of clustering, algorithm estimates load on current cluster, forms next level clusters, and adjusts it with network width. As comparisons showed, our algorithm outperforms other algorithms in terms of network lifetime and load balancing.

1. Introduction

Wireless Sensor Network (WSN) is composed of a large number of tiny sensor nodes that are densely deployed either inside the phenomenon or very close to it. These tiny sensor nodes, which consist of sensing, data processing, and communicating components, leverage the idea of sensor networks. Another unique feature of sensor networks is the cooperative effort of sensor nodes [1]. Node energy is an important factor to be considered in this network. Since, sensor nodes are equipped with small energy reserves and it is not possible to recharge energy of sensor nodes in many applications.

A great deal of research has directed to energy efficiency in sensor network. One of such research groups is clustering protocol. Clustering protocols deal with organizing nodes into clusters, selecting one node as cluster head (CH) in every cluster, sending data to Base station(BS) through CH. Clustering has been shown to improve network lifetime by aggregating data, reducing communication overhead and organizing routing paths efficiently. In [2], it is shown that clustering and data aggregation at least double the network lifetime.

Data transmission mode to BS can be single hop [3] or multihop [2] that is named inter-cluster communication in clustering protocols [4]. In single hop data transmission every CH sends data directly to BS. This mode is not energy efficient and it is impossible to transmit data for further CH from BS because of sensor capability and wireless channel property. Multihop mode maintains data transmission to BS through neighboring cluster heads (CHs). It is shown [3] that this type of transmission mode is more energy efficient than single hop mode. However, both of the modes cause problem of imbalance of energy consumption. In single hop mode CHs further than other CHs from BS consumes more energy and, in consequence, these CHs die earlier. The same problem occurs in relay CHs in multihop mode because of more burdens on relay CHs than other CHs. Network lifetime and functionality is significantly decreased by energy imbalance consumption problem. Network lifetime is one of the primary metric for evaluating the performance of a sensor network and it is defined either death of first node or some percentage of nodes depending on functionality of network [6]. Hence, energy consumption balancing is crucial while keeping energy efficiency.

In this paper, we propose and evaluate distributed clustering with load balancing (DCLB) for forming clusters efficiently and balancing load in inter-cluster communication. Cluster size (range) is crucial in terms of energy efficiency and balancing load in multihop communication of CHs. Since, it determines data volume and communication distance to the next CH. Hence, our algorithm considers the communication distance and data volume in every step of clustering to avoid energy inefficiency and to keep balanced load on clusters. It forms clusters with different sizes in every step, as shown in Fig. 1. The algorithm consists of two steps: Estimation of cluster ranges by BS and clustering process steps. Initially, BS determines the maximum and minimum range of clusters R_{max} and R_{min} which are important for energy efficiency and setting up target clusters efficiently. To determine cluster range in each level, previous CH's energy consumption to transmit data to the current CH and data volume is considered to estimate next level

978-0-7695-3507-4/08 $25.00 © 2008 IEEE
DOI 10.1109/CSIE.2009.816

cluster range. That is energy consumption of current CH to send its and previous CH's data should be smaller than previous CH's energy consumption. By this way cluster ranges in each level will be determined. To start clustering, BS sends the message which includes information about ranges of clusters approximate positions of candidate CHs and nodes in self organizing manner set up clusters. To further balance load and consume energy efficiently, we determine weight of every cluster *W* which CH exchanges with their neighbor to use to forward data toward BS.

2. Related Work

Clustering is considered one of the promising techniques for WSN. Hence, there are many works related to clustering with different objectives, assumptions and properties. We present some representative of relevant and basic research work.

Low Energy Adaptive Clustering Hierarchy (LEACH) is one of the earliest and most popular clustering algorithms. It forms clusters in distributed way, where nodes make decisions independently [3]. A node decides to be CH with a probability p, and broadcasts its decision. Every node chooses CH based on the least communication energy cost. CH is changed periodically based on random number "T" between 0 and 1 chosen by each node. A node becomes CH for the current round if number "T" is less than some threshold. Since, the decision is probabilistic which does not consider residual energy of node, there is probability that a node with low energy can be selected as CH. When this node dies, whole cell will loose its functionality. Moreover, it is assumed that CH sends data directly to BS. This is not always a realistic since wireless channel property makes this assumption unrealistic and CH is usual sensor node. In addition, it may not be applicable in large scale networks.

Another basic and popular clustering algorithm for WSN is Hybrid Energy-Efficient Distributed Clustering (HEED) [3]. In HEED, CHs are picked based on energy and communication cost. Only sensors that have a high residual energy can become CH. The algorithm consists of three phases to select CHs and form clusters. The phases are initialization, repetition, and finalization. In initialization, phase algorithm sets percentage of CH among sensor nodes. Every sensor node estimates its CH_{prob} - probability of being CH. CH_{prob} should not be below certain threshold p_{min}. In repetition phase, nodes which could not find its CH, try to find by iteration or announce themselves CHs. In finalization, phase every sensor node decides its status. It either picks the least cost CH or announces itself as CH. In HEED, inter-cluster communication based on multi hop where the shortest path tree routing is used. CHs near to BS deplete their energy much faster than other CH which causes to die much earlier of these clusters. Network lifetime can be affected seriously as a consequence.

To balance load on CHs in Multihop inter cluster communication, different schemes are proposed. The idea behind those schemes is that CHs take turn to send a data to BS in chain of CHs so that load distributed equally on CHs. In [5], authors proposed cluster-based even energy dissipation protocol (EDDP). EDDP presents schemes to balance load on CH and set up multiple chains of CH with different objectives in these chains. The scheme is proposed to balance load in a given chain of CHs with an equal distance among them. The chain with n CHs is divided into n rounds and chain can be divided into m sub-chains. If we assume that the chain is divided into two sub-chains, then in each round r, only r packets are relayed via node 1, and the remaining $n-r$ packets take a short cut from node $r+1$. The furthest CHs from BS consume energy rapidly. Direct communication is not practical and energy inefficient as we mentioned before. In addition, after sometime, nodes' energy level is reduced until some certain level of energy, then node can not support consuming energy rapidly. Even though nodes are alive and they have some energy but they can not participate in network operation because of lack of enough energy to direct communication with BS. The problem will be more complicated when this scheme is applied to chain of CHs where distance between CHs is different.

In [7], Energy-Efficient Unequal Clustering algorithm is proposed. This work is similar to our algorithm in the aspect of unequal clustering. Initially, several tentative CHs with the same probability T are selected. Each tentative CH broadcasts in range R^0_{comp} about its residual energy and competition range R_{comp} which is a function of node's distance to the BS. R^0_{comp} is the predefined maximum competition range. Then, every tentative CH competes to be CH if there one more tentative CH in its competition range R_{comp}. The node has higher residual energy will be CH and in case of tie node with smaller ID will win and be CH. The distance between candidate CHs is not considered which causes to be overlapped clusters sometimes. As a consequence not targeted clusters to be formed which affects load balancing. Also, authors introduced threshold TD_MAX in multihop forwarding model. If a node's distance to the BS is smaller than TD_MAX threshold, it transmits its data to the base station directly; otherwise it should find a relay node which can forward its data to the base station. In our opinion, this threshold is introduced to mitigate load on clusters near BS. However, its serious disadvantage is that it makes free clusters near BS from load. Clusters before that threshold distance are burden because of most amount of data and direct communication to BS. As a consequence, energy is consumed inefficiently and network lifetime is decreased.

3. Algorithm Description

3.1. System Model

Consider a sensor network consisting of a set of sensors dispersed in network field. In this network, the sensing tasks periodic data reporting. We assumed followings about properties of system model:

– The nodes are stationary in the network. This assumption is typical for WSN;
– BS is outside of network area and is stationary;
– Capabilities of sensor nodes are equal in terms of sensing, processing and communication;
– Nodes have fixed number of transmission level, that is, nodes can control radio transmission power and have a directional antenna;
– Nodes are uniformly distributed over the network area. Many large-scale sensor networks such as environment monitoring sensors dropped from aircraft have this property [1] [12];
– Nodes are able to estimate distance between them and BS. Precise inter-node distance estimations have been demonstrated using ultrasound in system proposed in [8], the MIT Crickets [9] and in the Medusa MK-2 node [10]. In the radio domain, ultra-wide-band ranging systems such as the one offered by Ubisense [13] have already demonstrated accurate distance measurements with small sensor from factors that will be suitable for sensor networks. In [11], authors mentioned that techniques like Received Signal Strength Indicator and Angle of Arrival give results whose accuracy satisfies the requirements of most applications;
– Network area is rectangular shape with N x M. Our solution can be easily extended to other network shapes.

3.2. Distributed Clustering Algorithm with Load Balancing

We propose distributed clustering with load balancing (DCLB) for energy efficiency and balanced inter cluster communication. The idea is to set up clusters in descending order in terms of BS and allocate them equally in horizontal as shown in Fig.1.

Initially, BS determines maximum and minimum R_{max} and R_{min} cluster range based on node density, limited range, and physical environment. R_{max} and R_{min} cluster range is important for energy and clustering mechanism efficiency.

Clusters with R_{max} are set up on the top of network that is furthest part from BS. The range can be adjustable according to a network width by slightly changing:

$$M/R_{max} = k \,. \quad (1)$$

M and k are width of network and number of clusters in the first layer that is in horizontal.

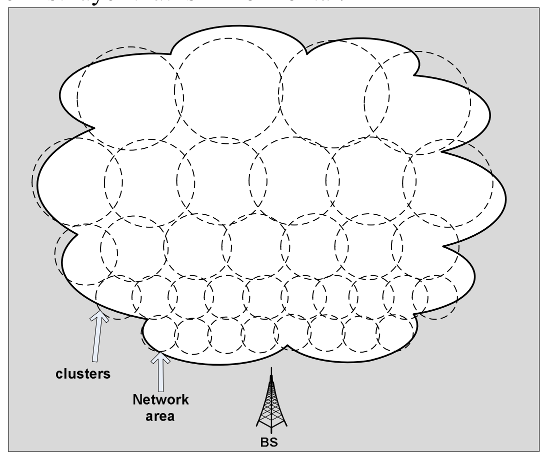

Fig.1. Distributed clustering with load balancing.

To make k integer number, R_{max} can be changed slightly. Next layer clusters' range could be determined as follows:

$$R_{max} - R_{min} = R_{max-1} \quad (2)$$

To achieve load balancing inter cluster communication and estimate next layer cluster range, BS first estimates energy consumption of CH of cluster range with R_{max} for transmission its data to CH of cluster range with R_{max-1} as follows [3]:

$$E\,(R_{max}+R_{max-1}) = l_{Rmax}\,(E_{elec} + E_{amp}\ (R_{max}+R_{max-1})^2) \quad (3)$$

where $E_{elec}l$ = 50nJ/bit, $E_{amp}l$ = 100pJ/bit/m^2, l is packet size, $R_{max}+R_{max-1}$ is distance between CHs, we consider that CHs are in the center of clusters and CH sends data straight to next CH. Packet size l_{Rmax} can be determined by following:

$$l_{Rmax} = \lambda \text{ x } a \text{ x } ((R_{max})^2 \text{x } \pi \text{ x } n) \ / (N \text{ x } M)^2 \quad (4)$$

where a is data of a node, it is same with all nodes, λ *is* aggregation coefficient, $(R_{max})^2 \text{x } \pi \text{ x } n / (N \text{ x } M)^2$ is number of nodes in cluster R_{max}, n is number of nodes in network.

Energy consumption of CH_{Rmax} that is CH of cluster range with R_{max} for sending data to CH of the cluster range with R_{max-1} can be used to find R_{max-2} such that load is balanced on CH_{Rmax-1} :

$$E\,(R_{max}+R_{max-1}) \ \geq (l_{Rmax} + l_{Rmax-1}) \text{ x } ((E_{elec} + E_{amp} \text{ x} (R_{max-1}+R_{max-2})^{2}) + E_{rec} \text{x } l_{Rmax}$$
$$R_{max-2} \geq R_{min} \quad (5)$$

The above inequality expresses balanced energy consumption of CH_{Rmax-1} with CH_{Rmax} for receiving data from CH_{Rmax} and sending all data to the distance R_{max-2}

which is the range of next layer clusters. By simply solving above quadratic inequality we can determine the cluster range in every layer until $R_{max-n} \leq R_{min}$. Then other side of network is adjusted with R_{min} range if there is space yet.

After finishing estimation cluster ranges in each layer, BS estimates approximate positions of CHs in each layer and their distances to the BS to form efficient clustering mechanism.

Before starting clustering process, in the first stage, BS broadcasts "Hello" message to all nodes [7] so that nodes can compute their distances to BS. In the second stage BS broadcasts over the network Form_Mes which includes exact and approximate distances Ex_Dist and App_Dist to BS of candidate CHs, range of a cluster. We introduce here not only exact distance from BS to network where nodes located but also approximate distances to guarantee clustering all part of network. Each node which received Form_Mes checks its distance with Ex_Dist and App_Dist and decides to be CH. If a node's distance to BS is same with Ex_Dist, then it directly broadcasts in related range that it is CH; if the distance belongs to the App_Dist then CH broadcasts with random back off time that it is CH. If nodes' distance matches neither Ex_Dist nor App_Dist, then it waits for CH announcement. By this way, clusters are set up.

To further balance load on CHs, we introduce W weight of clusters. A CH can exchange with their neighbors' weight information periodically so that load is distributed equally.

$$W_m = \frac{\sum_{i=1}^{n} R_i^E}{E_{CH_j} + \frac{1}{n}\sum_{i=1}^{n} E_i}$$

where Wm is weight of cluster m, R_i – residual energy of node i, E_{CHj} is energy consumption of CH_j in one round and Ei is average energy consumption of nodes in cluster m in one round.

4. Performance Evaluation

To evaluate performance of DLBC, we compare it with EEDP [5] and EEUC [7] algorithms which are presented in section 2. We evaluate the performance of the algorithms through numerical analysis in terms of energy efficiency, average lifetime, and energy consumption variance in CHs. To perform analysis, we take a sample chain of clusters according to each algorithm in 220m x 200m network with 400 nodes. Each node has 0.5J energy. A chain contains four clusters in EEDP and DCLB and five in EEUC. BS is located in the right side of network and distance between network and BS is 10 m. R_{max} and R_{min} is 40 m and 10 m. EEDP and EEUC algorithm parameters are similar as shown in authors' papers. Firstly, we estimated energy consumption of CHs in 32 rounds in each algorithm with parameters mentioned above. As Fig. 2 shows, EDDP consumes the highest energy because, each round one of CHs, it should send data directly to BS. It is burden especially for further CH as shown Fig.4 also. Next column shows the energy consumption in our algorithm which is little higher than EEUC. Cluster range is bigger than that in EUCC that means longer distances between CHs. Therefore energy consumption is higher but it can be tuned optimally to reserve energy by changing R_{max}. As bigger cluster range the more nodes it includes. As consequence gathered and transmitted data also will be more as shown Fig.3.

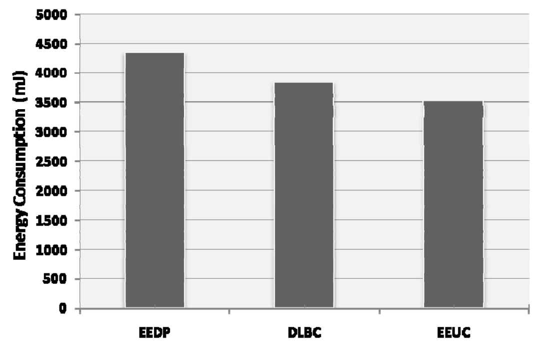

Fig.2. Energy consumption of CHs in algorithms.

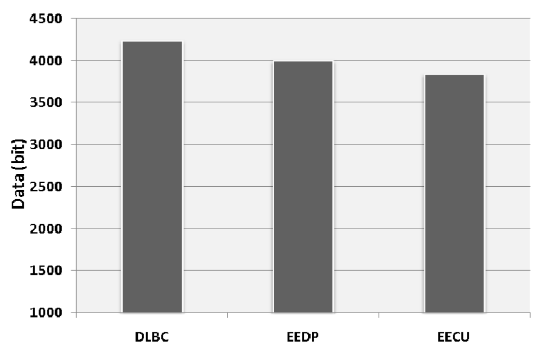

Fig.3. Transmitted data by CHs in algorithms.

Next, we evaluate algorithms in terms of energy consumption variance in CHs. Energy consumption variance influences network lifetime directly. Because, dissipation of energy equally over network while keeping energy efficiency provides long network lifetime. In Fig. 4, first column in each group of columns shows the energy consumption of CHs in EEDP. If energy consumption is less or equal in CHs near BS than other CHs in multihop like DCLB and EEUC, then it can achieve some load balancing. However, the case is different in EEDP Energy consumption should be equal in this algorithm. The second column in each group of columns shows EEUC algorithm performance and it generates one more cluster than other algorithms in given space. As we can see the difference is very high especially between CH1 and CH2. It is effect of TD_MAX threshold which restricts to some certain distance that nodes after that distance should send data directly to BS and nodes before that distance on burden as shown in the Fig 4. It influences network lifetime significantly as it is shown in Fig 5. Our algorithm outperforms both of two algorithms. Energy consumption is more stable and in increasing trend which is important in Multihop mode to load balancing. Since, it can control distance between CHs and consider relay data volume.

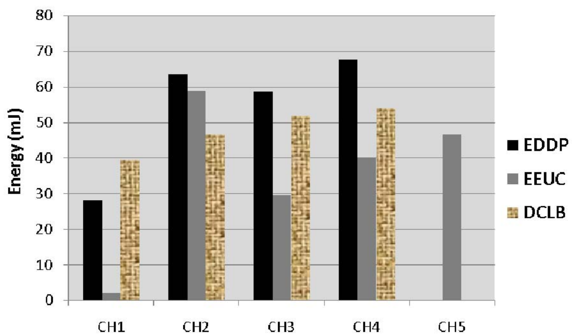

Fig.4. Energy consumption variance in CHs.

Fig. 5 shows the average lifetime of a chain which consists of four clusters in EEDP and DCLB and five clusters in EEUC. The second cluster died fully in 201 rounds in EEUC; second cluster died fully in 536 rounds in DCLB, fourth is died first fully EEDP in 636 rounds which is furthest cluster from BS. However, after 8 rounds, the other cluster died fully in EEDP and in DCLB next cluster died after 102 round. That is load balancing is

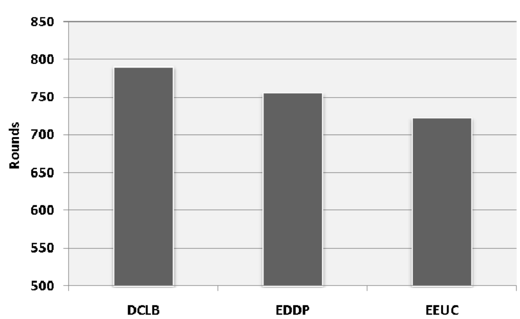

Fig.5. Average lifetime of CHs.

almost the same in EEDP and DCLB. However, life time of in EEDP is shorter than that of ours. In EEUC's, after dying second cluster, the performance is degraded because of TD_MAX threshold. Because, second cluster considered as before TD_MAX threshold distance so it is most loaded in terms of data volume and communication distance.

5. Conclusion

We have proposed distributed clustering with load balancing (DCLB) for energy efficiency and balanced inter-cluster communication. Clustering algorithm estimates cluster ranges considering distance and data value so that load is balanced on CHs. To achieve energy efficiency and to set up targeting clusters, R_{max} and R_{min} is introduced and they can be tuned to optimize energy efficiency. As evaluation results showed, it is more energy efficient than other algorithms and almost the same as EEDP in terms of energy consumption balancing. As a consequence, it provides longer lifetime.

6. References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: A Survey", Computer Networks (Elsevier) Journal, Vol. 38, No. 4, pp. 393-422, 2002.

[2] Younis and S. Fahmy, "Distributed Clustering in Ad Hoc Sensor Networks: A Hybrid, Energy-Efficient Approach," Proc. IEEE INFOCOM, Hong Kong, Mar. 2004;

[3] W. R. Heinzeman, A. Chandrakasan, and H. Balakrishnan, "Energy efficient communication protocol for wireless microsensor networks." In Proc. of HICSS, Jan. 2000.

[4] V. Mhatre and C. Rosenberg, "Design Guidelines for Wireless Sensor Networks: Communication, Clustering and Aggregation", Ad Hoc Networks, vol. 2, no. 1, pp. 45-63, 2004.

[5] Devendar Mandalat, Fei Dait, Xiaojiang Dutand Chao Yout, "Load Balance and Energy Efficient Data Gathering in WSN", In Proc. of IEEE MASS, Oct. 2006.

[6] Yunxia Chen, Qing Zhao, "On the lifetime of wireless sensor networks" Communications Letters, IEEE vol. 9, no. 11, pp. 976 – 978, Nov. 2005 .

[7] Chengfa Li, Mao Ye, Guihai Chen, Jie Wu, "An energy-efficient unequal clustering mechanism for wireless sensor networks", In Proc. of IEEE MASS, pp. – 604, Nov. 2005.

[8] A. Harter and A. Hopper, "A New Location Technique for the Active Office," IEEE Personal Communication, vol. 4, No. 5, pp. 42-47, October 1997.

[9] N. B. Priyantha, A. Chakraborty, H. Balakrishnan, "The Cricket Location Support System," Proceedings of 6th ACM Mobicom, Boston, MA, August 2000.

[10] A. Savvides, H. Park and M. B. Srivastava, "The n-hop Multilateration Primitive for Node Localization Problems," In Proc. of MNA, 8, pp. 443- 451, 2003.

[11] Michele Battelli, Stefano Basagni, "Localization for Wireless Sensor Networks: Protocols and Perspectives", In Proc. of CCECE 2007, pp 1074-1077, April 2007.

[12] F. Zhao and L. Guibas, Wireless Sensor Networks: An Information Processing Approach. Morgan Kaufmann, 2004.

[13] Ubisense website, http://www.ubisense.net.