Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I
IMECS 2009, March 18 - 20, 2009, Hong Kong

Centralized and Distributed Clustering Methods for Energy Efficient Wireless Sensor Networks

Noritaka Shigei*, Hiromi Miyajima*, Hiroki Morishita*, Michiharu Maeda†

Abstract— **The wireless sensor network (WSN) technology is a key component for ubiquitous computing. In order to achieve the long term operation of WSNs, communication protocols based on clustering have been extensively studied such as LEACH, ACE and HEED. In this paper, we propose two types of clustering methods for WSNs. The first type, which is based on centralized management, employs vector quantization (VQ) for effective clustering. The second type, which is performed in a distributed distributed fashion, takes into account remaining battery level and node density. The proposed methods are compared with conventional methods LEACH, HEED and ANTCLUST. The effectiveness of the proposed methods are demonstrated by numerical simulation.**

Keywords: wireless sensor network, energy efficient, clustering, vector quantization

1 Introduction

The wireless sensor network (WSN) technology is a key component for ubiquitous computing. A WSN consists of a large number of sensor nodes as shown in Fig.1. Each sensor node senses environmental conditions such as temperature, pressure and light and sends the sensed data to a base station (BS), which is a long way off in general. Since the sensor nodes are powered by limited power batteries, in order to prolong the life time of the network, low energy consumption is important for sensor nodes. In general, radio communication consumes the most amount of energy, which is proportional to the data size and proportional to the square or the fourth power of the distance. In order to reduce the energy consumption, a clustering and data aggregation approach has been extensively studied[7]. In this approach, sensor nodes are divided into clusters, and for each cluster, one representative node, which called cluster head (CH), aggregates all the data within the cluster and sends the data to BS. Since only CH nodes need long distance transmission, the other nodes save the energy consumption.

In order to manage effectively clusters and CHs, distributed clustering methods have been proposed such as LEACH, HEED, ACE and ANTCLUST[2, 3, 4, 6]. LEACH, which is the most popular method, guarantees that every nodes evenly become CHs but does not take into account battery level and the interrelationship among nodes[2]. HEED, ACE and ANTCLUST achive better performance than LEACH by taking into account battery level, communication cost, node density, etc. However, they need additional inter-node communications for determining clusters and CHs.

In this paper, we propose two types of clustering methods with less communication overhead for clustering. The first type, which is based on centralized management, employs vector quantization (VQ)[1] for effective clustering. In the centralized method, BS determines clusters and CHs according to battery level and node location. The second type, which is performed in a distributed autonomous fashion, takes into account battery level and node density. In the distributed method, clustering is performed by the interaction among proximity nodes. The proposed methods are compared with conventional methods LEACH, HEED and ANTCLUST. The effectiveness of the proposed methods are demonstrated by numerical simulation.

2 Wireless Sensor Network

2.1 WSN model

This section describes the wireless sensor network (WSN) model considered in this paper[2, 3, 4, 6]. The WSN model consists of N sensor nodes and one base station (BS) node as shown in Fig.1. All sensor nodes are identical and are assumed to have the following functions and features: 1) sensing environmental factors such as temperature, pressure, and light, 2) data processing by low-power micro-controller, 3) radio communication, and 4) powered by a limited life battery. The BS node is assumed to have an unlimited power source, processing power, and storage capacity. The data sensed by sensor nodes are sent to the BS node over the radio, and a user can access the data via the BS node. In this WSN application, the clock synchronization of sensor nodes is an important issue. Because the time at which a data was sensed is important, which requires low clock skew among all the sensor nodes. We assume that the low clock skew

*Dept. of Electrical and Electronics Engineering, Kagoshima University, 1-21-40 Korimoto, Kagoshima 890-0065, Japan (corresponding author N. Shigei to provide email: shigei@eee.kagoshima-u.ac.jp)
†Fukuoka Institute of Technology, Fukuoka 811-0295, Japan

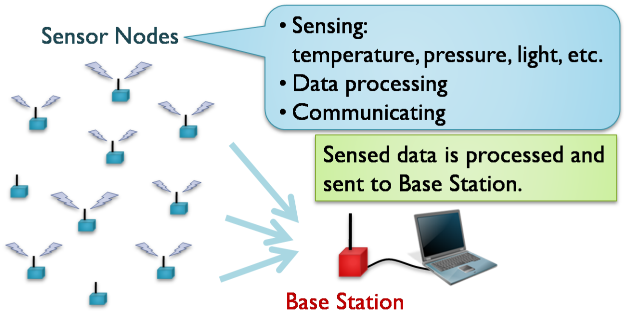


Figure 1: The concept of wireless sensor network.

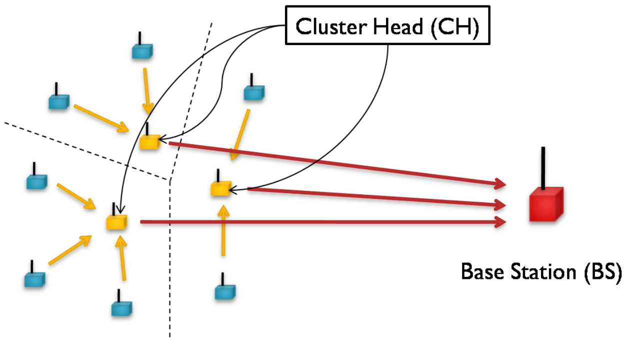


Figure 2: The concept of the clustering approach for WSN.

requirement is guaranteed by using a clock synchronization method[5].

The radio communication consumes more energy than the data processing on a sensor node. We assume the following energy consumption model for radio communication. The transmission of a k-bit message with transmission range d meters consumes $E_{\mathrm{T}}(k,d)$ of energy.

$$E_{\mathrm{T}}(k,d) = \begin{cases} k(E_{\mathrm{elec}} + \varepsilon_{\mathrm{fs}} d^2) & \text{for } d \le d_0 \\ k(E_{\mathrm{elec}} + \varepsilon_{\mathrm{mp}} d^4) & \text{for } d > d_0, \end{cases} \quad (1)$$

where E_{elec} is the electronics energy, and $\varepsilon_{\mathrm{fs}}$ and $\varepsilon_{\mathrm{mp}}$ are the amplifier energy factors for free space and multipath fading channel models, respectively. The reception of a k-bit message consumes $E_{\mathrm{R}}(k)$ of energy.

$$E_{\mathrm{R}}(k) = k \cdot E_{\mathrm{elec}} \quad (2)$$

2.2 Clustering Approach for WSN

In order to save the energy consumption of WSN, a clustering approach for WSN as shown in Fig.2 has been considered. In the approach, N sensor nodes are divided into clusters, and each cluster has a representative sensor node called cluster head (CH). Each non-CH sensor node sends the sensed data to the CH node in its own cluster, instead of to BS. Each CH node aggregates the received data into smaller size and sends it to BS. This approach has the following advantages: 1) non-CH sensor nodes can save the energy consumption because the nodes can avoid long-distance communication and have only to send data to its own CH being nearby and 2) the amount of data to be sent to BS can be reduced, which also saves the energy consumption.

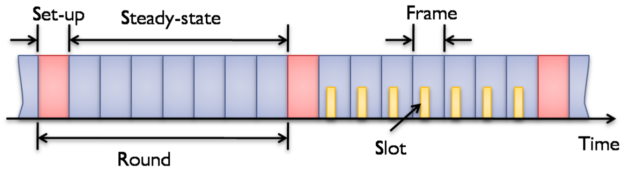


Figure 3: The operating cycle in clustering methods.

The operating cycle of clustering methods is shown in Fig.3. Each round consists of consecutive frames. The first frame is for set-up, and the others are for steady-state. In the set-up frame, CH nodes and clusters are determined based on the used clustering algorithm, and each CH assigns a non-CH node to a slot in order to create time-division multiple-access (TDMA) schedule. In the steady-state frames, each non-CH node sends data to CH at the assigned slot in TDMA fashion, and CHs fuse (compress) the received data and send it to BS.

In order to decide CHs and clusters, clustering algorithms such as LEACH and HEED have been proposed[2, 3]. In LEACH, CHs are determined in a distributed autonomous fashion. At each round l, each node v independently decides to be a CH with probability $P_v(t)$ if the node v has not been a CH in the most recent $(l \bmod (N/k))$ rounds.

$$P_v(l) = \frac{k}{N - k(l \bmod \frac{N}{k})}, \quad (3)$$

where k is the average number of CHs for each round. This means that each node becomes CH at least once every N/k rounds. However, LEACH does not take into account battery level and node distribution.

3 The Proposed Methods

More effective clustering methods than LEACH have been proposed such as HEED, ACE and ANTCLUST[3, 4, 6]. However, they need additional inter-node communications for clustering. In this section, we propose two types of methods with less inter-node communication for clustering. The first method is a centralized method, and the second is a distributed method.

3.1 The Centralized Method

In this method, the BS node manages the clustering by utilizing a vector quantization (VQ) technique. The traditional VQ process approximates the distribution of the large set of vectors $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_\nu\}$ by using a small set of vectors $W = \{\boldsymbol{w}, \cdots, \boldsymbol{w}_\kappa\}$, where $\nu \gg \kappa$ in general, $\boldsymbol{x} \in X$ is called *input vector*, and $\boldsymbol{w} \in W$ is called *weights* or *codebook vectors*[1]. In VQ, an input vector $\boldsymbol{x} \in X$ is approximated by the nearest weight $\boldsymbol{w}_{\iota(\boldsymbol{x},W)}$, where $||\boldsymbol{x} - \boldsymbol{w}_{\iota(\boldsymbol{x},W)}|| = \min_{\boldsymbol{w}\in W} ||\boldsymbol{x} - \boldsymbol{w}||$. An adaptive VQ algorithm such as LBG and K-means trains the weight set so as to minimize the approximation error $E = \frac{1}{|X|}\sum_{\boldsymbol{x}\in X} ||\boldsymbol{x} - \boldsymbol{w}_{\iota(\boldsymbol{x},W)}||^2$. Assuming that each $\boldsymbol{w} \in W$ and each $\boldsymbol{x} \in X$ are correspond to a CH node and a non-CH node, respectively. This assignment of CHs and clusters will minimize the total energy consumption of intra-cluster communication because the energy consumption of transmission is proportional to the squared distance as shown in Eq.(1).

The direct application of VQ will not work because for the same X it always gives a same assignment, which brings specific nodes down quickly. In order to successfully apply VQ to the clustering of WSN, we propose a VQ method utilizing remaining battery level information. The algorithm is presented as follows:

Algorithm Centralized Clustering
Input:
$A = \{$All active nodes$\}$.
$X = \{$The coordinate of node v $\boldsymbol{x}_v | v \in A\}$.
Step 1: Each active node $v \in A$ sends its coordinates $\boldsymbol{x}_v$ and its remaining battery level e_v to BS.
Step 2: BS performs the following VQ procedure.
(2-1) Initialize the weight set $W = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k\}$ by random numbers and $t \leftarrow 0$.
(2-2) Select randomly a node $v \in A$ with probability p_v.

$$p_v = \frac{e_v}{\sum_{u\in A} e_u} \tag{4}$$

(2-3) Find the nearest $\boldsymbol{w}_k \in W$ to $\boldsymbol{x}_v$, where $||\boldsymbol{x}_v - \boldsymbol{w}_k|| = \min_{\boldsymbol{w}\in W} ||\boldsymbol{x}_v - \boldsymbol{w}||$.
(2-4) Update $\boldsymbol{w}_k$ as follows.

$$\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k - \alpha(\boldsymbol{w}_k - \boldsymbol{x}_v). \tag{5}$$

(2-5) $t \leftarrow t+1$. If $t = T_{\max}$ then go to Step 3. Otherwise go to (2-2).
Step 3: For each $k \in \{1, \cdots, K\}$, k-th CH is assigned to the nearest node $v \in A$ to $\boldsymbol{w}_k \in W$, where $||\boldsymbol{x}_v - \boldsymbol{w}_k|| = \min_{u\in A} ||\boldsymbol{x}_u - \boldsymbol{w}_k||$. Let $C = \{v \in A$ is CH.$\}$ be the set of CH nodes. Let $c(k) \in C$ denote the k-th CH node.
Step 4: Each $v \in A \setminus C$ is assigned to k-th cluster whose CH is the nearest CH from v, that is $||\boldsymbol{x}_v - \boldsymbol{x}_{c(k)}|| = \min_{j\in\{1,\cdots,K\}} ||\boldsymbol{x}_v - \boldsymbol{x}_{c(j)}||$.
Step 5: BS broadcasts the decided CH and cluster assignments. □

3.2 The Distributed Method

The second proposed method is performed in a distributed autonomous fashion. In the method, the role of BS is just to receive sensed data from CHs. There exist some conventional distributed methods such as LEACH, HEED and ANTCLUST. The distinguished features of the proposed method are as follows: 1) aware of remaining battery power level, 2) aware of node density, and 3) a small communication overhead in the clustering process. LEACH has a small communication overhead but is not aware of remaining battery level and node density. HEED and ANTCLUST are aware of remaining battery level and node density but their communication overheads are large.

Unlike our centralized version in the previous section, in our distributed version, each sensor node broadcasts its own existence within its proximity. Based on the proximity information and their own battery level, the CH nodes and the clusters are autonomously determined among sensor nodes. In the process, every nodes broadcast some messages at most twice within a small range of R_{inf} or R_{cnd} radius. The proposed algorithm consists of four phases which performed in a setup frame. In l-th round, each node begins each phase $q \in \{1, 2, 3, 4\}$ at a specific time $(l-1)T_{\text{rnd}} + \sum_{j=1}^{q-1} T_j$, where $T_{\text{rnd}} = \sum_{j=1}^{4} T_j + T_{\text{ss}}$ and T_{ss} is the period of the stead state. The algorithm is presented as follows:

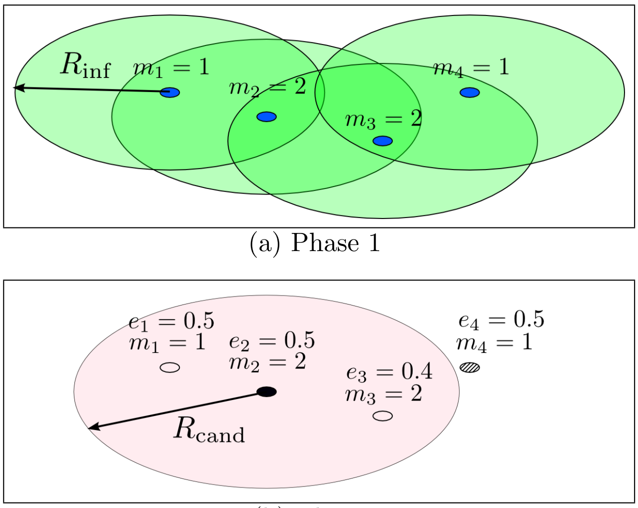
Algorithm Distributed Clustering
Phase 1: All active nodes broadcast their node IDs within R_{inf} meters radius. All nodes count how many IDs are received. Let m_v be the counted number for node v.
Phase 2: Each node v broadcasts its candidacy for CH within R_{cnd} meters radius in descending order of the following evaluation function

$$f(m_v, e_v) = m_v e_v^4, \tag{6}$$

provided that a node that receives a candidacy before its broadcasting does not broadcast its candidacy.
Phase 3: The nodes that broadcast the candidacy in Phase 2 become CHs. The other nodes becomes non-CHs. The non-CH nodes send *intentions of participating* to the nearest candidate for CH.
Phase 4: Each CH node creates a TDMA schedule and send it to the non-CH nodes as the registration approval. □

Fig.4 shows examples of Phases 1 and 2. In Fig.4.(a) for Phase 1, all the active nodes broadcast their node IDs, and nodes 1, 2, 3 and 4 recognize one, two, two and one neighboring nodes, respectively. In Fig.4.(b) for Phase 2, node 2 has the maximum $m_v \cdot e_v^4$ and broadcasts its candidacy for CH in the first place. As a result, nodes 1 and 3 give up their candidacy, and node 4 will broadcast its candidacy in time.

In the algorithm, the way to execute the phase 2 is not obvious. Let us explain the phase 2 in detail. The period of "phase 2" T_2 consists of the candidacy period $T_{2\text{c}}$ and the post-margin $T_{2\text{m}}$, that is $T_2 = T_{2\text{c}} + T_{2\text{m}}$. The margin



(a) Phase 1

(b) Phase 2

Figure 4: Examples of phases 1 and 2.

$T_{2\text{m}}$ is inserted to prevent the effect of unavoidable clock skew and communication delay. In l-th round, each node v broadcasts its candidacy for CH at time $t_{\text{cnd}}(l, m_v, e_v)$, where

$$t_{\text{cnd}}(l, m, e) = (l-1)T_{\text{rnd}} + T_{2\text{c}} \left(1 - \frac{m \cdot e^4}{\alpha(l)}\right), \qquad (7)$$

and $\alpha(l)$ is a decreasing function with round l such that $\alpha(l) < \max_{v'} m'_v \cdot e'^4_v$. If $m_v \cdot e_v = \max_{v'} m'_v \cdot e'^4_v$ then $t_{\text{cnd}}(l, m_v, e_v) > t_{\text{cnd}}(l, m'_v, e'_v)$ for any $v' \neq v$, that is the first broadcast of candidacy is performed by the node v. If $m_v \cdot e^4_v = \min_{v'} m'_v \cdot e'^4_v$ then $t_{\text{cnd}}(l, m_v, e_v) < t_{\text{cnd}}(l, m'_v, e'_v)$ for any $v' \neq v$, that is the last broadcast of candidacy is performed by the node v. Further, if $m_v \cdot e_v > m'^4_v \cdot e'^4_v$ then the node v broadcasts earlier than the node v'.

4 Numerical Simulation

In this section, the effectiveness of the proposed methods is demonstrated by numerical simulation. The proposed methods are compared with the conventional methods LEACH, HEED and ANTCLUST.

In the simulation, N sensor nodes are randomly distributed in the square region of size 100 m × 100 m and the base station is 75 meters away from the center of a side as shown in Fig.5. The parameters used in the simulation is summarized in Table 1. The simulation is performed for $N = 100$, 300 and 1000.

The simulation results for $N = 100$, 300 and 1000 are shown in Figs. 6, 7 and 8, respectively. In the graphs, “Centralized” and “Distributed” are our proposed methods in subsections 3.1 and 3.2, respectively. For $N = 100$ and $N = 300$, our distributed method is the best and our centralized method is the second best. The difference between our proposed methods for $N = 300$ is small than for $N = 100$. For $N = 1000$, our centralized method is the best and our distributed method is the second best. For any N, our proposed methods achive better performance than the conventional methods LEACH, HEED and ANTCLUST. Our simulation result shows that our centralized and distributed methods are suitable for larger and smaller numbers of sensor nodes, respectively.

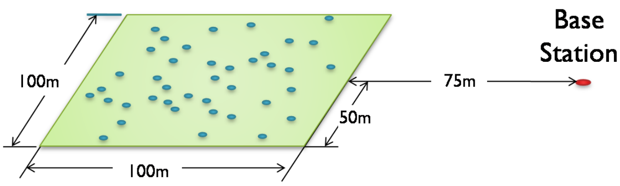


Figure 5: The node arrangement in the simulation.

Table 1: Parameters used in the simulation.

For energy model	
d_0	75 m
E_{elec}	50 nJ/bit
E_{fusion}	5 nJ/bit
ε_{fs}	100 pJ/bit/m^2
ε_{mp}	1.3 fJ/bit/m^4
Initial battery level	0.5 Joule
Energy for data aggregation	5 nJ/bit/signal
For packet model	
Data packet size	800 bit
Broadcast packet size	200 bit
Packet header size	200 bit
For distributed method	
R_{inf}	20 meters
R_{cnd}	55 meters

5 Conclusions and Future Work

In this paper, we proposed two types of clustering methods for WSNs. The first type, which is based on centralized management, employs vector quantization (VQ) for effective clustering. The second type, which is performed in a distributed autonomous fashion, takes into account remaining battery level and node density. The effectiveness of the proposed methods were demonstrated in the numerical simulation. In the simulation, our proposed methods prolong the network lifetime longer than the conventional methods LEACH, HEED and ANTCLUST. Further, our simulation results show that the centralized method and the distributed method are suitable for larger and smaller sensor nodes, respectively. Our future works are theoretical analysis of the proposed methods, further improvement of prolonging performance, consideration on other models such as WSN models with solar cell, and evaluation on a WSN testbed.

ISBN: 978-988-17012-2-0

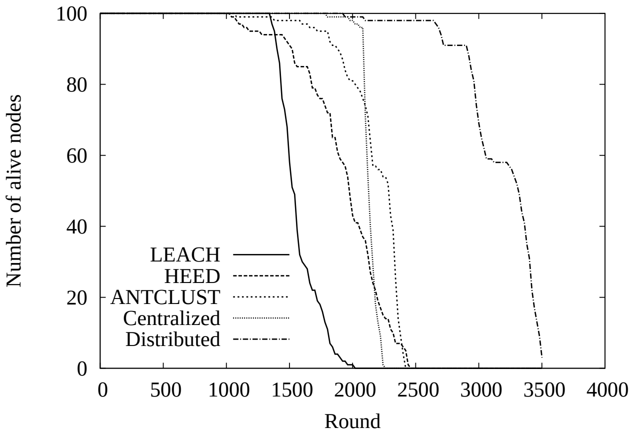


Figure 6: The number of alive nodes versus round for $N = 100$.

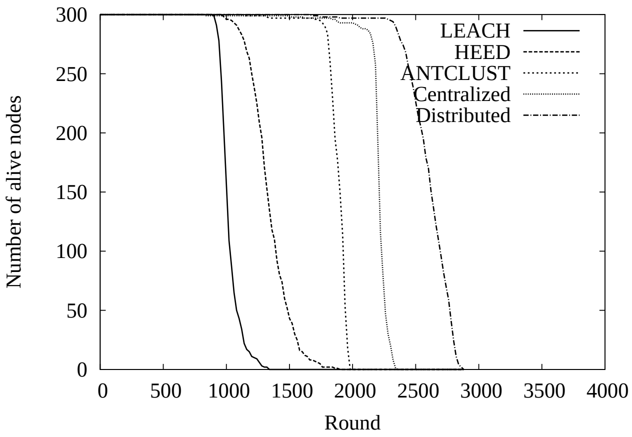


Figure 7: The number of alive nodes versus round for $N = 300$.

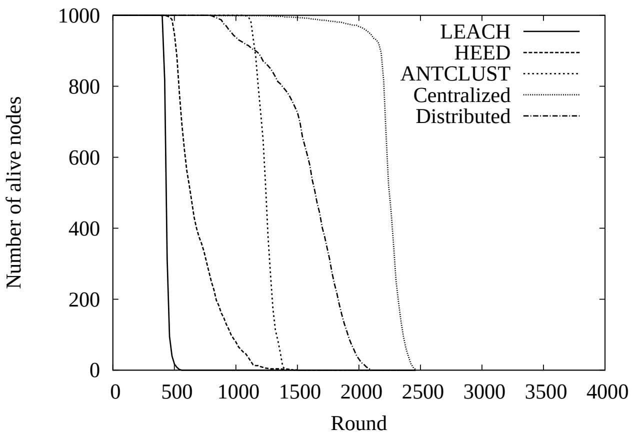


Figure 8: The number of alive nodes versus round for $N = 1000$.

Acknowledgements

This work is supported by Grant-in-Aid for Scientific Research (C) (No.20500070) of Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] Linde, Y., Buzo, A., Gray, R.M., "An Algorithm for Vector Quantizer Design." *IEEE Trans on Communications*, V28, pp.84-95, 1980.

[2] Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H., "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE Trans on Wireless Communications*, V1, N4, pp. 660-670, 2002.

[3] Younis, O., Fahmy, S., "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks," *IEEE Trans on Mobile Computing*, V3, N4, pp. 366-379, 2004.

[4] Chan, H., Perrig, A., "ACE: An Emergent Algorithm for Highly Uniform Cluster Formation," *Proc. 1st Euro Workshop Sensor Networks*, pp.154-71, 2004.

[5] Sundararaman, B., Buy, U., Kshemkalyani, A.D., "Clock Synchronization for Wireless Sensor Networks: a Survey," *Ad Hoc Networks*, V3, pp.281-323, 2005.

[6] Kamimura, J., Wakamiya, N., Murata, M., "A Distributed Clustering Method for Energy-Efficient Data Gathering in Sensor Networks," *International Journal of Wireless and Mobile Computing*, V1, I2, pp.113-120, 2006.

[7] Abbasi, A.A., Younis, M., "A Survey on Clustering Algorithms for Wireless Sensor Networks," *Computer Communications*, V30, pp.2826-2841, 2007.