❒   2371

K-means clustering-based WSN protocol for energy efficiency improvement

Jong-Yong Lee[1], Daesung Lee[2]
[1]Ingenium College of Liberal Arts, Kwang Woon University, Republic of Korea
[2]Department of Computer Engineering, Catholic University of Pusan, Republic of Korea

Article Info	ABSTRACT
	Since it is very difficult to replace or recharge the batteries of the sensor nodes in the wireless sensor network (WSN), efficient use of the batteries of the sensor nodes is a very important issue. This has a deep relationship with the lifetime of the network. If the node's energy is exhausted, the node is no longer available. If a certain number of nodes (50% or 80%) in a network consume energy completely, the whole network will not work. Therefore, various protocols have been proposed to maintain the network for a long time by minimizing energy consumption. In recent years, a protocol using a K-means clustering algorithm, one of machine learning techniques, has been proposed. A KCED protocol is proposed in consideration of residual energy of a node, a cluster center, and a distance to a base station in order to improve a problem of a protocol using K-average gung zipper algorithm such as cluster center consideration.

Corresponding Author:

Daesung Lee
Department of Computer Engineering
Catholic University of Pusan,
57, Oryundae-ro, Geumjeong-gu, Busan, Republic of Korea
Email: dslee@cup.ac.kr

1.   INTRODUCTION

Sensor nodes in WSN [1-10] are often deployed in large quantities, mainly where it is difficult for people to access. Therefore, it is very important to use the limited energy of sensor nodes efficiently because it is difficult to replace or recharge batteries in sensor nodes. For these reasons, one of the most important considerations in the WSN design is to minimize energy consumption at each node to increase the energy efficiency of the network. If the wireless sensor node consumes all of its energy, the node is no longer usable, and if more than a certain (50% or 80%) of the node in the network consumes all of its energy, the network becomes inoperable. Thus, various protocols were proposed to minimize the energy consumption of the nodes and maintain the network for a long time [11-15]. The LEACH protocol is a hierarchical clustering algorithm for energy efficiency, where the cluster head is elected through a probability threshold. However, it is not always possible to ensure that an optimal cluster is formed. In addition, actual clustering may result in very few or too many nodes in the cluster, and the cluster head election may result in data transfer failures as well as first node dead (FND). The problem of nodes with less residual energy being elected as cluster heads has been improved by taking into account energy terms in the probability threshold. In other words, the energy-consideration LEACH protocol was proposed to minimize the election of nodes with less energy remaining as cluster heads. By multiplying the remaining amount of energy in the nodes by the cluster head election threshold, the lower the residual energy, the lower the probability of cluster head election. However,

the energy-consideration LEACH protocol still cannot guarantee optimal cluster formation or even node density per cluster.

Wireless sensor protocols using the K-means clustering algorithm [16-22] do not form a cluster after electing the cluster head, but do the cluster configuration first. This technique has the advantage of having a uniform cluster configuration, with most of the member nodes in the cluster present uniformly. And this method, after the cluster configuration, elected as the cluster head either a node with a lot of residual energy or a node close to the cluster center point. However, there is a problem with nodes far away from base stations becoming cluster heads, or with the same nodes being cluster heads in succession. This causes FND to occur quickly. When using the K-means clustering algorithm, the member nodes within the cluster are as uniform as possible. This can improve unbalanced cluster configurations, such as the LEACH protocol. However, because the clustering process is performed repeatedly moving the center point to the final cluster finalization, it has the disadvantage of taking longer time to cluster than traditional hierarchical algorithms. In addition, for the K-means clustering-based protocol, only the remaining energy of the node or node close to the cluster's center point was considered when electing the cluster head, but not the transmission distance to the base station that consumes the most energy. In this paper, we are going to improve the problem of K-means clustering to increase the energy efficiency of WSN. The proposed protocol takes into account the residual energy of the node and the distance to the base station when electing the cluster head. To take into account the residual energy of the node and the distance to the base station, the cluster head-elected Score algorithm is used for each node.

2. RESEARCH
2.1. LEACH protocol

The low-energy adaptive clustering hierarchy (LEACH) protocol [23-24] is a typical clustering-based protocol proposed by Wendy B. Heinzelman. The LEACH protocol consists of a set-up phase and a steady-state phase. In the set-up phase, the cluster head is randomly elected by the probability threshold equation and the cluster configuration is performed. The probability threshold T(n) used to elect cluster heads for node n is expressed in (1) and has a value between 0 and 1.

$$T(n) = \begin{cases} \frac{p}{1-p\left(r\,mod\frac{1}{p}\right)} & if \quad n \in G \\ 0 & otherwise \end{cases} \tag{1}$$

In (1), p is the probability of electing a cluster head, r is the current round, and G is the set of nodes that were not elected as cluster heads until the previous round. Each node generates a random number between 0 and 1, comparing it to equation (1) and then the node is elected as the cluster head when it has a value less than the probability threshold $T(n)$. Since cluster heads were elected by equation (1) the cluster heads broadcast advertising messages to surrounding nodes in sensor field, and the normal nodes receiving advertising messages join as clusters of cluster heads with the largest signal strength to form clusters. When the cluster configuration is complete, the cluster head creates and assigns a time division multiple access (TDMA) schedule, which specifies the time each node must transmit, depending on the number of member nodes in the cluster. During the steady-state phase, the data is sent according to the TDMA schedule assigned by the cluster head. The cluster head completes the steady-state phase by aggregation the data received from the member node and sending it to the base station in a code division multiple access (CDMA). The cycle completed from these set-up phase to the steady-state phase is called Round.

The LEACH protocol improved the problem of cluster head election for specific nodes of existing clustering-based protocols, by electing all nodes with remaining energy as cluster heads once and for all. However, by using only the probability threshold when electing a cluster head, the network has a short lifetime span due to problems such as the election of nodes with insufficient residual energy. To improve this, various protocols that modified the probability threshold were proposed.

2.2. Energy-consideration LEACH protocol

In the case of (1) used to elect the cluster head of the LEACH protocol, elect the cluster head without considering the residual energy of the node. Thus, even if the node has less residual energy, it can be elected as a cluster head. To improve this M. J. Handy [25] proposed the following modified threshold (2):

$$T(n)_{new} = T(n) \times \frac{E_{current}}{E_{max}} \tag{2}$$

In (2) E_max is the maximum (initial) energy of the node, and E_current is the residual energy of the node. For the cluster head election threshold, the value will be between 0 and 1. The closer the value is 1, the higher the chance of cluster head election, and the closer the value is to zero, the lower the chance of cluster head election. M. J. Handy's proposed threshold for residual energy is to multiply the value of conventional (1) by the residual energy ratio of the node, and the lower the residual energy, the lower the probability of cluster head election. Thus, minimizing the election of nodes with less residual energy as cluster heads, results in increasing network lifetime. In Gupta's fuzzy logic, the chance value for all nodes are calculated for every round. When the chance value calculation is completed, the cluster heads are elected in ascending order of the chance value. The centrality of the node is the sum of the distances from the node located within a certain range r to the node A as shown in Figure 1. The range is given by (2).

3. PROPOSE METHOD

Using the K-means clustering algorithm in the WSN protocol results in the member nodes in the cluster being as uniform as possible, but it takes longer to cluster than traditional hierarchical algorithms. In addition, cluster head election has the problem of continuous cluster head election on the same node, with only considering the residual energy of the node or node that is mostly close to the cluster's center point. This results in shortening the lifetime span of the network. And when electing a cluster head, not considering the distance to the node and base station, the overall energy efficiency of the network can be reduced. This paper proposes a protocol that increases the energy efficiency of the network while improving these problems. To improve the disadvantage of taking a long time to configure a cluster, the proposed protocol limited the cluster configuration point to the first round at which the system was initialized and the next round at which additional sensor nodes were consumed with all the remaining energy. This can improve the time-consuming problem because cluster configurations do not occur at every round. By taking into account the distance from the cluster center point and the residual energy of the node when electing the cluster head, the problem of not taking into account the transmission distance to a base station that consumes a lot of energy. To solve these problems, the cluster head was elected in this paper considering the residual energy of the node and the distance to the cluster's center point or base station.

In general, if only the residual energy of the nodes in the cluster is considered, the cluster head candidate has elected the node with the most residual energy. However, considering the distance to the base station and the remaining energy, simply electing the cluster head according to the ranking decision can lead to an energy-efficient problem. For example, for any two nodes A and B, if the residual energy in node A is greater than the residual energy in node B, and node A is farther than node B, the residual energy is taken into account and node A is elected as the cluster head. In this case, because of the large energy consumption due to the distance to the base station, the energy in node A is consumed quickly, thereby reducing the lifetime span of the entire network. To improve this problem, the remaining energy and distance Score computations for nodes in the cluster are proposed and elect the cluster head by using a Score operation on member nodes that satisfy $E_{residual} > AVERAGE(E_{cluster})$, which has more residual energy than the average residual energy for all nodes in the cluster. This has the advantage of reducing the amount of computations consumed in elections. The Score is divided into $Score_{BS}$ based on the base station distance and $Score_{CC}$ based on the cluster's center point, respectively, as defined in (3) and (4).

$$Score_{BS} = \frac{E_{residual}}{E_{init}} - \frac{d_{toBS}}{MAX(d_{toBS})} \qquad (3)$$

$$Score_{CC} = \frac{E_{residual}}{E_{init}} - \frac{d_{toCC}}{MAX(d_{toCC})} \qquad (4)$$

Here, $E_{residual}$ means the current residual energy of the node, E_{init} represents the initial energy of the node, d_{toBS} refers to the distance to the node and base station, and d_{toCC} refers to the distance to the node and cluster's center point. The ratio of residual energy to the initial energy, the first term of (3) and (4) means the relative size (normalization value) of the residual energy. This value has a value between 0 and 1 and decreases to zero as the round progresses. And the second term of (3) and (4) has a value between 0 and 1 as the normalized terms for the distance. That is, the larger the first term and the smaller the second term, the greater the probability of being elected as the cluster head.

In other words, if the Score values for (3) and (4) are large, they are elected as the cluster head. By using the proposed technique, the Score operation on nodes in a cluster can determine which node is close to the base station or cluster's center point while the node has a high residual energy. Once the $Score_{BS}$ and $Score_{CC}$ have been computed, elect a node with values MAX($Score_{BS}$) and MAX($Score_{CC}$) as the cluster head candidate. If a node in a cluster has a maximum Score value for both the base station and the cluster's

center point, it is elected as the cluster head. However, if the node with MAX($Score_{BS}$) values and the node with MAX($Score_{CC}$) values are different, the cluster head was determined by calculating: To predict the energy consumption of the two cluster head candidate nodes, the sum of the distance from the member nodes in the cluster and the distance from the base station, i.e. the total transmission distance, were computed. The total transmission distance *TotalDistance (i)* for i-node can be obtained using (5).

$$TotalDistance(i) = d_{Node_i toBS} + \sum_{j=1}^{totalMN} d_{Node_j toNode_i} \tag{5}$$

Here, the first term is the distance to the i-node and base station (d_(Node_i toBS)), and the second term is the sum of the distances between the i-node and the rest of the nodes in the cluster. Subsequently, a candidate node with a short total transmission distance was finally elected as the cluster head. If multiple nodes with the same Score values were elected as cluster head candidates, the node closest to the base station was elected for $Score_{BS}$ and the node closest to the cluster's center point for $Score_{CC}$ as the cluster head candidate. In addition, if the values of the cluster head candidates were the same, the shorter the distance to the base station, the less energy was consumed, so the node with MAX($Score_{BS}$) was elected as the cluster head. Once the cluster head was elected, the data was collected and sent to the base station in the same way as the LEACH protocol. The flowchart of the propose protocol works as shown in Figure 1 and Figure 2.
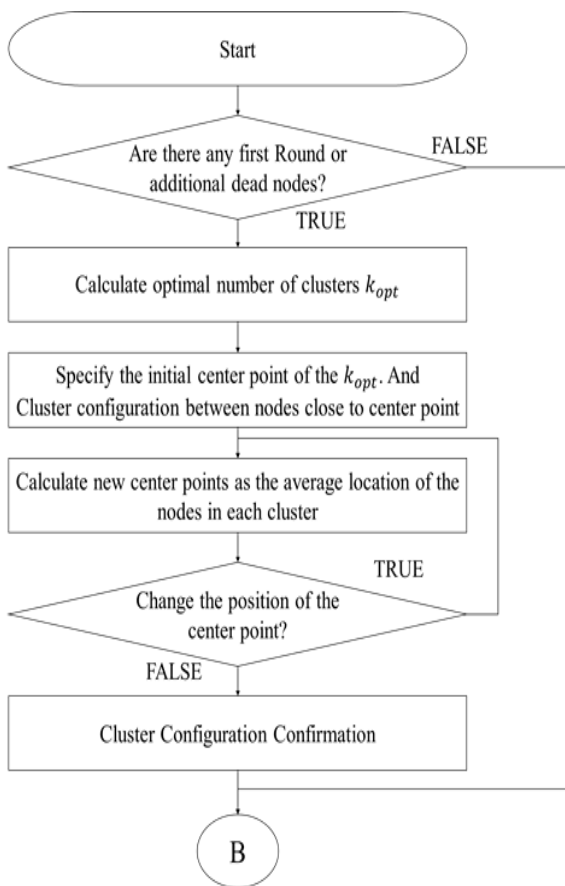

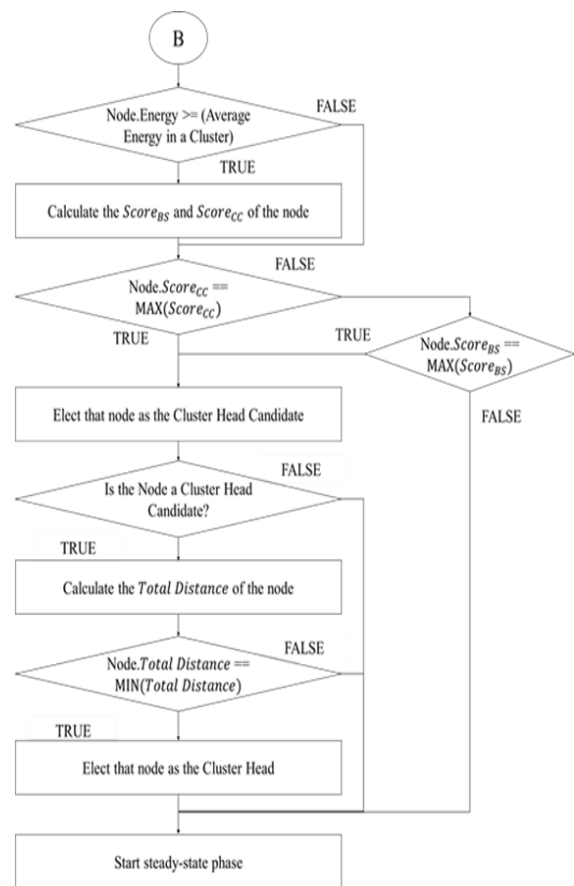Figure 1. Flowchart of propose protocol's cluster head election

Figure 2. Flowchart of propose protocol's cluster head election

Figure 3 shows the pseudo-code of the proposed protocol. The two to seven lines of code are the process of checking for additional dead nodes compared to the previous round, if any, for clustering. Other than that, existing clusters are maintained. The 9-36 line of code is the process of calculating $Score_{BS}$ and $Score_{cc}$ of nodes in each cluster, and then electing nodes with $MAX(Score_{BS})$ and $MAX(Score_{cc})$ as cluster candidates. The 37-42 line is the process of computing the total transmission distance of the cluster candidates and then electing a node with a short distance as the final cluster head.

```
1    // for every round
2    prevDeadNodeCount = -1
3    deadNodeCount = CheckDeadNode()
4    IF NOT prevDeadNodeCount == deadNodeCount
5    StartClustering()
6    prevDeadNodeCount = deadNodeCount
7    END
8    FOR i = 1 to k
9    MaxBSNode = Cluster(i).MemberNode(1)
10   MaxBSScore = Node(MaxNode).CurrentEnergy / InitEnergy –
     Node(MaxNode).DistanceToBS /
     MAX(Node.DistanceToBS)
11   MaxCCNode = Cluster(i).MemberNode(1)
12   MaxCCScore = Node(MaxNode).CurrentEnergy / InitEnergy –
     Node(MaxNode).DistanceToCC /
     MAX(Node.DistanceToCC)
13   FOR j = 1 to Cluster(i).MemberNodeCount

14   IF Node(j).CurrentEnergy >= Cluster(i).AverageEnergy
15   Node(j).ScoreBS = Node(j).CurrentEnergy/InitEnergy–Node(j).DistanceToBS /
     MAX(Node.DistanceToBS)
     Node(j).ScoreCC = Node(j).CurrentEnergy/InitEnergy–Node(j).DistanceToCC /
     MAX(Node.DistanceToCC)
16   IF Node(j).ScoreBS > MaxScoreBS
     MaxBSScore = Node(j).ScoreBS
     MaxBSNode = j
17   ELSEIF Node(j).ScoreBS == MaxBSScore
18   IF Node(j).DistanceToBS < Node(MaxBSNode).DistanceToBS
19   MaxBSScore = Node(j).ScoreBS
20   MaxBSNode = j
21   END
     END

22
23   IF Node(j).ScoreCC > MaxCCScore
24   MaxCCScore = Node(j).ScoreCC
25   MaxCCNode = j
     ELSEIF Node(j).ScoreCC == MaxCCScore
26   IF Node(j).DistanceToCC < Node(MaxCCNode).DistanceToCC
27   MaxCCScore = Node(j).ScoreCC
28   MaxCCNode = j
29   END
30   END
     END
31   END

32
33   IF TotalTransDistance(MaxBSNode) <= TotalTransDistance(MaxCCNode)
34   MaxNode = MaxBSNode
35   ELSE
36   MaxNode = MaxCCNode
     END
37   SetClusterHead(MaxNode)
     END
38
```

Figure 3. Pseudo-code for the propose protocol (KCED)

4. SIMULATION AND RESULT

To verify the energy efficiency of the proposed protocol, it was compared with the LEACH protocol, the energy-consideration LEACH protocol, and simulation with MATLAB. Assumptions about the sensor field for the simulation are as follows: In the simulation environment, all the same sensor nodes are constructed, and once deployed, the nodes are not moved. It was then considered that all sensor nodes had the same initial energy and that base stations were located outside the sensor field. The simulation parameters are defined as shown in Table 1. The results of comparing network lifetime using the simulation parameters are shown in Figure 4 and Table 2. It can be seen that the proposed protocol is an improvement of 111% in FND criterion compared to the LEACH protocol and an 80% improvement in FND criterion over the energy-considerations LEACH protocol.

Table 1. Simulation energy model

Parameter	Value
Number of Sensor Node	100
Size of Sensor Field	200 x 200
Location of Base Station	(100, 300)
Initial Energy of Sensor Node	0.5J
E_{DA}	5nJ/bit/signal
E_{elec}	50nJ/bit
ϵ_{fs}	10pJ/bit/m^2
ϵ_{mp}	0.0013pJ/bit/m^4

Table 2. Simulation result: network lifetime comparison

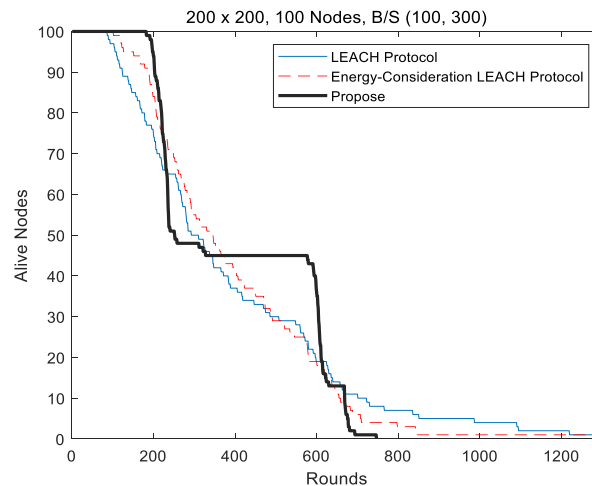| Protocol	FND	80% Alive
LEACH Protocol	86	171
Energy-Consideration LEACH protocol	101	206
Propose Protocol	182	219
Improvement rate	111% ▲ vs.LEACH protocol	28% ▲ vs.LEACH protocol
	80% ▲ vs. Energy-Consideration LEACH Protocol	6% ▲ vs.Energy-Consideration LEACH Protocol

Figure 4. Simulation result: alive node graph by round

5. CONCLUSION

In this paper, the cluster configuration is efficient in wireless sensor network, and the method of cluster head election is proposed to improve the network lifetime. For efficient cluster configuration, the K-means clustering algorithm was used. This can make the member nodes as uniformly part of the cluster as possible. The proposed protocol reduced the computations by requiring cluster configuration to be refreshed only in the first round and the next round when the dead node occurred. An the cluster head was then elected through the $Score_{BS}$ and $Score_{CC}$ operations. By using $Score_{BS}$, It can elect nodes that are as close to base stations as possible while still have more energy, and by using $Score_{CC}$, it can elect nodes that are as close to the cluster center as possible while still have more energy remaining. After electing these two candidates, by finalizing a node with a low total transmission distance as a cluster head, we were able to elect the appropriate cluster head and simulation results showed that the network lifetime would improve.

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] J. Y. Lee, "Energy Improvement of WSN Using The Stochastic Cluster Head Selection," *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, vol. 15, no. 1, pp. 125-129, 2015.

[3] S. Subedi, and S. Lee, "A New Technique for Localization Using the Nearest Anchor-Centroid Pair Based on LQI Sphere in WSN," *Journal of Information and Communication Convergence Engineering*, vol. 16, no. 1, pp. 6-11, 2018.

[4] Brar, Gurbinder Singh, et al., "Energy efficient direction-based PDORP routing protocol for WSN," *IEEE access*, vol. 4, pp. 3182-3194, 2016.

[5] Wang, Zhong-xun, et al., "A clustering WSN routing protocol based on node energy and multipath," *Cluster Computing*, vol. 22, no. 3, pp. 5811-5823, 2019.

[6] Anzola, John, et al., "A clustering WSN routing protocol based on kd tree algorithm," *Sensors*, vol. 18, no. 9, p. 2899, 2018.

[7] Wang, Quan, et al., "An energy-efficient compressive sensing-based clustering routing protocol for WSNs," *IEEE Sensors Journal*, vol. 19, no. 10, pp. 3950-3960, 2019.

[8]   Mohamed, Reem E., et al., "Energy efficient collaborative proactive routing protocol for wireless sensor network," *Computer Networks*, vol. 142, pp. 154-167, 2018.

[9]   Kumar, Rajeev, and Kumar Dilip, "Multi-objective fractional artificial bee colony algorithm to energy aware routing protocol in wireless sensor network," *Wireless Networks*, vol. 22, no. 5, pp. 1461-1474, 2016.

[10] Singh O., Rishiwal V., Yadav M., "Energy trends of routing protocols for H-WSN," *2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall)*, Bareilly, 2016, pp. 1-4.

[11] Jamal N. Al-Karaki, Ahmed E. Kamal, "Routing techniques in wireless sensor networks a survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6-28, 2004.

[12] Zhao Cheng, Mark Perillo, Wendi B. Heinzelman, "General network lifetime and cost models for evaluating sensor network deployment strategies," *IEEE Transactions on Mobile Computing*, vol. 7, no. 4, pp. 484-497, 2008

[13] Joanna Kulik, Wendi Heinzelman, Hari Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, pp. 169-185, 2002.

[14] G. Sanjeev, et al., "Cross layer architecture based mobile WSN routing protocol for inter-vehicular communication," *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT)*, Ghaziabad, 2017, pp. 1-7.

[15] Kushal B. Y., and Chitra M., "Cluster based routing protocol to prolong network lifetime through mobile sink in WSN," *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, 2016, pp. 1287-1291.

[16] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881-892, 2002.

[17] Razzaq Madiha, Ningombam Devarani Devi, Shin Seokjoo, "Energy efficient K-means clustering-based routing protocol for WSN using optimal packet size," *2018 International Conference on Information Networking (ICOIN)*, Chiang Mai, 2018, pp. 632-635.

[18] Sharma, Deepak Kumar, et al., "kROp: k-Means clustering based routing protocol for opportunistic networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 4, pp. 1289-1306, 2019.

[19] Ray Anindita, and D. E. Debashis, "Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wireless Sensor Systems*, vol. 6, no. 6, pp. 181-191, 2016.

[20] D. R. Renuga, and Sethukarasi T., "Energy Efficient Enhanced K-Means Cluster-based Routing Protocol for WSN," *Energy*, vol. 6, no. 11, 2019.

[21] S. Basma, "Energy optimization in wireless sensor networks using a hybrid k-means pso clustering algorithm," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 4, pp. 2679-2695, 2016.

[22] Mahboub Aziz, et al., "Energy-efficient hybrid k-means algorithm for clustered wireless sensor networks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 4, pp. 2054-2060, 2017.

[23] Razaque Abdul, et al., "P-LEACH: Energy efficient routing protocol for Wireless Sensor Networks," *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, Farmingdale, NY, 2016, pp. 1-5.

[24] S. Sunil Kumar, K. Prabhat, S. Jyoti Prakash, "A survey on successors of LEACH protocol," *IEEE Access*, vol. 5, pp. 4298-4328, 2017.

[25] M. J. Handy, M. Haase, and D. Timmermann, "Low energy adaptive clustering hierarchy with deterministic cluster-head selection," *4th International Workshop on Mobile and Wireless Communications Network*, Stockholm, Sweden, 2002, pp. 368-372.

BIOGRAPHIES OF AUTHORS

Jong-Yong Lee He received the B.S. degree in Nuclear Engineering from HanYang University, Korea, in 1983, the M.S. and Ph.D degree in Electrical Engineering from KwangWoon University, Seoul, South Korea, in 1986 and 1992. His current research interests include nonlinear system analysis and control, feedback linearization, computer aided control, computer network, image fusion and WSN, Sensor Network.
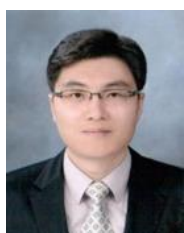
Daesung Lee He is a professor in the Department of Computer Engineering, Catholic University of Pusan, Korea. He received the B.S., M.S. and Ph.D. degrees from the Inha University, Korea, in 1999, 2001 and 2008, respectively, all in Electrical Engineering Computer Science & Engineering from Inha University. His research interests include security in network, convergence and operating system.