Received: January 26, 2022. Revised: March 15, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Clustering Optimization for Energy Consumption Problems in Wireless Sensor Networks using Modified K-Means++ Algorithm

Fransiska Sisilia Mukti[1]* **Allin Junikhah[1]** **Putu Mas Anggita Putra[2]** **Aryuanto Soetedjo[3]**
Awan Uji Krismanto[3]

[1]*Department of Informatics, Faculty of Technology & Design,*
Institut Teknologi dan Bisnis Asia Malang, Indonesia
[2]*Department of Informatics, Faculty of Mathematic and Sciences,*
Udayana University, Bali, Indonesia
[3]*Department of Electrical Engineering, Faculty of Industrial Technology,*
Institut Teknologi Nasional Malang, Indonesia
* Corresponding author's Email: ms.frans@asia.ac.id

Abstract: Wireless Sensor Network (WSN) became viable solution for smart technology for many industries nowadays. With abundant credentials, WSN still left a major constraint in energy consumption. A numerous of small and tiny sensors are equipped with limited resources bring the necessity to design a reliable WSN with efficient energy and network lifetime. Dividing nodes into clusters is believed as one of the efficient methods to cover these problems. But in fact, clustering process in basic routing protocols is not always guarantee an optimal cluster formation. Low Energy Adaptive Clustering Hierarchy (LEACH) is the most classical and famous routing protocol in WSN who offers clustering in ameliorating the nodes lifetime. Nevertheless, cluster imbalanced are often found in LEACH due to its cluster head (CH) random selection. For this, many studies integrated clustering algorithms into LEACH, such as k-Means. Actually, k-Means performs better clustering than basic LEACH, but we found that k-Means brings asymmetrical cluster formation which may leads empty cluster (in worst case). Due to these phenomena, we proposed an improvement clustering algorithm for LEACH using modified k-Means++ algorithm named Modified k-Means++ for LEACH (MKPP-LEACH) with three main considerations: computational declaration of first centroid, distance as additional parameter in determining the cluster centre, and take into account the residual energy in selecting CH. We simulated our proposed approach and compared it with basic LEACH and integrated k-Means for LEACH (KM-LEACH). The simulation demonstrates that MKPP-LEACH put much respect to energy efficiency with produce less energy and can keep most alive nodes almost 42,8% compared to LEACH and 57,2% compared to KM-LEACH.

Keywords: WSN, LEACH, k-Means, k-means++, Clustering optimization, Energy consumption.

1. Introduction

WSN was succeeded attract a lot of attention of researchers in the past decades, which has driven the evolution of Internet of things (IoT) [1]. Given credence more benefits compared to wired networks, WSNs claimed as appealing technology for smart infrastructure, like automation and control applications [2].

WSNs comprises of numerous small and tiny sensors, which are formed as self-organized and self-configured components [3]. They are responsible in monitoring the environment, recording, and sending the aggregated data to the collector (namely base station/BS) [4]. This technology is suitable for various smart systems such as renewable energy systems which is required to collect a large quantities of data [1].

As independent sensor nodes (SN), they are equipped with limited energy sources and processors. They also alimented by batteries (in most of the cases they are neither rechargeable nor replaceable) [5]. Furthermore, in most of WSN applications, SN are often deployed in tough spots or hazardous environments [6], where the replacement of components may be an impossible task [1, 7]. So, they must be able to self-organized and connect-effortless into their communication.

These limited characteristics bring the necessity to design a reliable system with efficient management in energy and network lifetime. The challenges for the researchers are providing the most energy-efficient method to extend the SN lifetime [8]. Hence, energy consumption still became the major constraint in WSNs [5].

The energy usually consumed in case of sensing, transmitting, or receiving packets and collecting the data. Mostly, redundancy data occurs and causes the SN energy depletion [5]. Thereby, route selection is a critical activity in overcoming the severe resource constraints in WSN [9].

Routing or route selection in WSNs is more challenging than other wireless networks [4]. Many studies are designed with aims to implement highly efficient routing protocols [9] in facing up the diversity of WSN topology formation. Inappropriate routing results in more energy consumption, so it is important to conceive what factors influenced the node's energy [4].

Routing protocols in WSNs are divided into three types, they are flat routing, hierarchical (cluster-based) routing and location-based routing [10]. Hierarchical routing architecture has proved to be extremely effective in solving the problem of excessive energy consumption [9].

The most classical and famous hierarchical or clustering routing protocol in WSN [5] is LEACH, proposed by Heinzelman [11]. LEACH ameliorates the lifetime of WSNs by decreasing the number of transmission packets via clusters formation [5] It integrates both concepts of energy efficient cluster-based routing and media access together [12]. In LEACH, the nodes in the network self-organise into different clusters, there is only cluster head of each cluster, and all the other nodes send their data to the head of the cluster [13]. Using clustering techniques, resource-constrained nodes do not need to send their data to gateways (sink) directly which can cause energy depletion, resource consumption inefficiency and interference [14].

However, it has numerous drawbacks, such as the random selection of CHs without considering either the distance or the current energy of nodes [5]. This scheme can easily lead to uneven energy consumption between network nodes, and shorten the network life [15]. Also, LEACH does not guarantee the position and number of CHs in each round. Formation of clusters in basic LEACH is random and leads to unequal distribution cluster in the network [16]. Another drawback is the single-hop used by CHs to communicate with the sink, which makes LEACH is not applicable for large networks [5].

A number of new protocols have been proposed with LEACH as benchmark; Kumar Singh et al [16] indicate that there are more than 60 extended versions of LEACH were deployed. Most studies considered that energy conservation (or efficiency) as the most common and important objective for WSN applications, so they focused on investigating the optimal clustering method for LEACH.

Clustering is believed as an efficient method indeed in improving network lifetime by decreasing the energy utilization [10], but it is not always guarantee that optimal cluster will be formed [17]. Therefore, many studies integrated clustering algorithms and schemes into WSN applications [11].

K-Means appears as clustering algorithm which is mostly used by many researchers [3, 18-21]. This algorithm belongs to unsupervised learning method for data clustering, introduced by MacQueen [3]. K-Means in WSN works when establish the cluster formation [19] and the algorithm ensures that SN will be uniformly distributed for each cluster. K-Means helps LEACH in mitigating the energy consumption for data transmission and prolong the network lifetime, compared to the conventional LEACH protocol.

Nevertheless, it takes a longer time in clustering formation than conventional LEACH. In addition, the cluster head (CH) election was considered based on its residual energy of the closest SN to the central point of each cluster. The distance between SN and BS is not taken into account when selecting the CH, for this reason, the overall energy efficiency cannot be known appropriately [17, 19]. Moreover, k-Means algorithm sets the initial centroid randomly and leads different kinds of clusters for each round. In worst case, it may bring out an empty cluster, even asymmetrical cluster formation [22].

Due to these drawbacks, we proposed an improvement clustering algorithm for LEACH protocol using modified k-Means++. We pay attention for the energy consumptions and unequal cluster formation based on basic k-Means for LEACH and resolved them using our proposed protocol. Some considerations to answer the problems are pointed as follows: (1) the initial centroid will be declared computationally (2) the distance between SN and BS is taken into account in determining the central point for each cluster to achieve more symmetrical cluster (3) put on the residual energy as additional parameters in selecting CH to avoid SN depletion.

This rest of the paper is organized as follows. Section 2 summarizes some related works on extended LEACH protocol based on clustering algorithm. In section 3, we describe the notation and some preliminaries. Section 4 proposes the detail of our approaches, which followed by analysis and performance evaluation of the proposed protocol in the section 5. Finally, conclusions are presented in the section 6.

2. Related works

A dynamic clustering using distributed k-means approach and silhouette method was presented at [23]. The clustering process is triggered again when the percentage of dead nodes reach 5 % of the alive nodes. The re-clustering process are calculated using silhouette method while k-means plays a role in creating the cluster. The simulation results that their proposed routing scheme achieves lower energy consumption and longer network lifetime compared to LEACH protocol. The authors did not consider about the trigger threshold for re-clustering process, which might lead more energy when the process is repeated periodically.

Another execution of k-means algorithm for LEACH protocol also described in [13]. They proposed a new algorithm called KDUCR which a combination of k-Means and Dijkstra algorithm. K-means is used to divide SN into k clusters and assign CH for the first round, while Dijkstra algorithm is used as shortest path counter between CH and BS. The results show that KDUCR performs energy consumption better than LEACH and LEACH-C protocols. K-means clustering algorithm is adopted by the sink node, and all of the clustering process is generated by sink node. This allows the reduction of energy of whole network, but actually causes the energy in the sink node to run out faster.

A new hybrid protocol called MDC-LEACH-K, which is a combination between LEACH-K approach and MDC was proposed in [24], to improve the LEACH protocol aggregation. Specifically, this protocol uses k-means clustering algorithm to reduce energy consumption in CH election phase. In addition, a mobile data collector (MDC) is used as an intermediate between CH and BS to further enhance the QoS criteria of WSN, to minimize time delays during data collection and to extend the lifespan of the network in WSN. This protocol achieves a significant energy gain by 296 % of residual energy compared to LEACH and 257 % to LEACH-KMeans protocol (LEACH-K). The authors set a fixed k values for simulation (k=10), which is indicate that there is no flexibility in this study, it is possible that another k values may provide better performance than 10.

In [25], energy efficient multi-hop routing clustering protocol (EEMRCP) based on fuzzy k-means and centralized mid-point algorithm (FKM-CMA) is proposed for network lifetime improvement. There are two main considerations for cluster head selection in proposed work, one is residual energy, and another one is Euclidean distance used in basic fuzzy k-means algorithm. Finally, multi-hop communication is performed for transmitting the packets from CHs to base station depending on the distance between them. But the simulation shows that this method is not suitable for large network based on its first death node.

A combination of k-means algorithm and slime mould algorithm (SMA) was presented in [15]. The proposed routing algorithm, called SMA-LEACH, is superior to other algorithms, namely PSO-LEACH, BA-LEACH, which using particle swarm optimization (PSO), bat algorithm (BA) to improve LEACH, respectively. Simulation analysis shows that the SMA-LEACH obviously reduces network energy consumption and extends the lifetime of WSNs. The two algorithms work independently (SMA for CH selection while k-means for clustering) and the discussion of the advantages of the method are only limited to the number of alive nodes. There is no further explanation regarding the energy consumption used.

The previous surveys mentioned above concentrated on aggregating k-means algorithm with other methods in order to achieve energy efficiency on LEACH through cluster formation techniques. Whereas the clustering process at k-means is strongly influenced by the initialization of k. On the contrary, this study presents a clustering optimization by taking into account of k initialization, distance and residual energy. Our proposed protocol suggests a complete clustering techniques and CH selection criteria to reduce energy consumption and improve the whole network lifetime.

3. Preliminaries

3.1 LEACH protocol

LEACH deployed SN into local clusters and divided the communication process into two rounds called set-up phase and steady-state phase. For each cluster, a CH will be assigned who is responsible in creating schedule for cluster members (CMs) while transmitting data. LEACH adopted time division multiple access (TDMA) method in making the

Received: January 26, 2022. Revised: March 15, 2022.

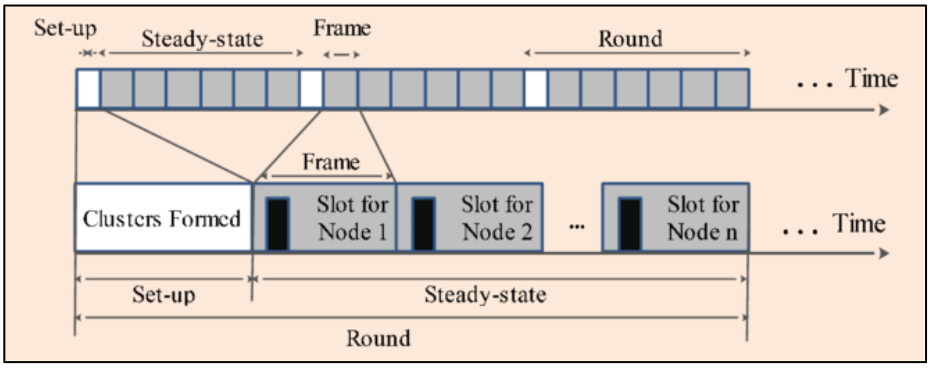


Figure. 1 Two phases operation in LEACH

schedules and informed it to the CMs. CH has in charge of collecting the data form CMs and doing some aggregation process in alleviating data redundancy before sent it to the BS. Relatively, CH works as relay node to help CMs in shortening the transmission distance in order to preserve their energies [26].The communication process under LEACH protocol is described as follows.

3.1.1. Set-up phase

Setup phase denotes the process of CH selection and cluster establishment. Every node has a chance being CH [4] and randomly self-elected as CH after generating random number between 0 and 1 [27]. If the selected value is larger than the threshold function which is defined in Eq. (1), then the node is selected as CH, otherwise, they will be a CM.

$$T(n) = \begin{cases} \frac{p}{1-p \times \left[r \bmod \left(\frac{1}{p}\right)\right]}, & n \in G \\ 0, & n \notin G \end{cases} \quad (1)$$

where p is the percentage of CH over the network, r is the number of rounds, and G is the set of nodes that not been elected as CH in $1/p$ rounds [6]. If a node selects itself as CH, they will make a broadcast message and inform this to its surround nodes. A CM may receive more than one broadcast message from different CHs, but they can decide which cluster to follow based on the distance and signal strength. Hereinafter, CM and CH will exchange their information to make a record which will be used by CH in creating TDMA schedules. This schedule will be broadcasted into all members then CMs will get their idle slots for data transmission. Then, they are ready to start the steady-state phase [26].

3.1.2. Steady-state phase

Steady state phase invents longer time than setup phase because it is time to every CM to forwards data to own CHs in their assigned time slots (based on TDMA schedules before) [27]. CH will collect the data and aggregate them firstly before sent to the BS, because of a CM may be sense similar data with another CM, so this aggregation process can diminish the unnecessary traffic in form of data redundancy. For this, CH has to keep up communication status at all times to receive data from all of the CM, which leads more energy consumption during this heavy tasks [26]. Fig. 1 depicts the communication process under LEACH protocol [28].

3.2 K-means algorithm

K-means is the simplest unsupervised learning algorithm that partitioning the data set into k cluster using euclidian distance [23]. The following steps in creating a cluster using k-means algorithm is describe below [21].

(1) Define the number of desired clusters
(2) Generate k to identify the cluster centres and the number of desired clusters
(3) Calculate the distance between nodes and cluster centres using Eq. (2).

$$Dist\ (x_1, x_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \quad (2)$$

(4) Assign nodes into closest centre based on its distance
(5) Calculate new cluster centre based on the mean value of all nodes in the respective cluster.
(6) Repeat step 2 using new cluster centre.
(7) Repeat step 4 if there is a change in node points, else stop the process.

For its implementation in LEACH protocol, there are some adjustments in setup phase explicitly in clustering process. Every node is represent as coordinate (*x*,*y*) and to fulfil this requirement, euclidian distance is modified in Eq. (3).

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \left| x_j - c_i \right|^2 \quad (3)$$

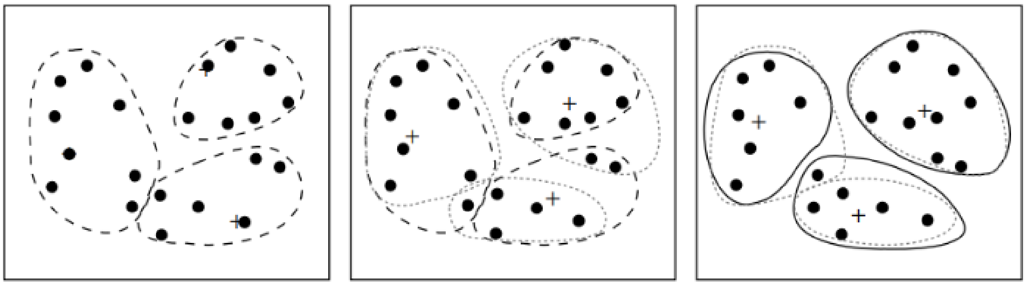


Figure. 2 Clustering objects using k-means method

k-Means Algorithm for LEACH Protocol
Inputs: *k*, *C*, *i* Outputs: a set of *k* clusters Initialize simulation parameter in creating WSN topology Deploy nodes in *n* * *n* unit 1. Input *k* as desired cluster numbers 2. Do *C* iteration to determine the initial centroid for each cluster and put *C* randomly. 3. For all nodes in network repeat below step (4-6) 4. Calculate the distance between nodes and *C* using Eq. (2) 5. Classify all nodes to the initial centroid based on closest distance 6. For all nodes in network repeat below step. **If** there is a value change of *C*, **Then** recalculate C_i using Eq. (3) **Else** *i* = cluster centroid.

Figure. 3 Pseudocode of k-means algorithm for LEACH

where J represents the distance between node and centroid, x_j as the *j* node to *i* cluster and c_i means the centroid for *i* cluster The k-means procedure for LEACH protocol is summarized in Fig. 3.

3.3 K-means++ algorithm

In the case of finding initial centroids, k-means clustering used randomization. The initial *k*-centroids were picked randomly from the data points. This randomization of picking *k*-centroids points results in the problem of initialization sensitivity. This problem tends to affect the final formed clusters. The final formed clusters depend on how initial centroids were picked.

There might be a way to resolve this problem by repeating the algorithm and initialization centroids in several times until getting the smaller intracluster distance and larger intercluster distance. But this process may take more energy consumption and times. Therefore a smart centroid initialization technique was introduced by Arthur and Vassilvitskii [29] called *k*-means++.

K-means++ presents to overcome the drawback in k-means, with ensuring a smarter initialization of the centroids and improves the quality of the clustering. Apart from initialization, the rest of the algorithm is the same as the standard k-means algorithm. The steps to follow for centroid initialization are described below [30-31].

(1) Randomly select the first centroid from the data points.
(2) For each data point (x_i) compute its distance (d_i) from the nearest, previously chosen centroid using Eq. (4).

$$d_i = max_{(j:1\to m)}\left|x_i - C_j\right|^2 \quad (4)$$

where d_i represents as the distance between x_i from the farthest centroid, while m as the number of centroids already picked.
(3) Make x_i as the new centroid who has the maximum probability proportionally to d_i
(4) Repeat steps 2 and 3 until *k* centroids have been sampled.

3.4 Energy dissipation model for WSN

The lifetime of a SN is usually measured by its energy level after a given network data transmission round. The energy level in a SN after each network data transmission phase is referred to as its residual energy. Given that the energy of a SN at the initial deployment is defined as ε_{in} , and the energy consumed by the SN after a particular data transmission round is ε_r, the residual energy of the SN ε_0 can be defined as Eq. (5).

$$\varepsilon_0 = \varepsilon_{in} - \sum_{r=0}^{R} \varepsilon_r \quad (5)$$

where $\varepsilon_0 = \varepsilon_{in}$ for *r* = 0, *r* is the current round, and *R* is the maximum number of rounds. From Eq. (5), when ε_0 = 0J, the SN has completely depleted its energy and cannot participate in any network activity [32].

Consider the simplified WSN energy dissipation model illustrated in Fig. 3 [47]. As shown in the figure, the transmitting SN consumes energy to drive its radio subsystem which includes the radio electronics, and power amplifier. The receiving SN also dissipates energy to drive its radio electronics. The distance *d* between the transmitting SN and receiving SN usually computed in the form of a distance metric specifies the channel model used by the power amplifier. If the distance *d* is greater than a set threshold d_0, the multipath model is assumed. Otherwise, the free space channel model is used for *d* < d_0. Therefore, the required energy for a SN to transmit a *q* bit message is defined in Eq. (6) [3], [32].

$$\varepsilon_{TX}(q,d) = \begin{cases} q \times \varepsilon_{elec} + q \times \varepsilon_f \times d^2, & d < d_0 \\ q \times \varepsilon_{elec} + q \times \varepsilon_m \times d^4, & d \geq d_0 \end{cases} \quad (6)$$

where ε_f and ε_m are the free space and multipath power loss respectively, ε_{elec} is the energy dissipated to drive the radio electronics, and d_0 is the transmission distance threshold expressed as Eq. (7) [32].

$$d_0 = \sqrt{\frac{\varepsilon_f}{\varepsilon_m}} \quad (7)$$

On the other hand, the energy consumed by the receiving SN is defined as [32].

$$\varepsilon_{RX} = q \times \varepsilon_{elec} \quad (8)$$

4. Proposed MKMPP-LEACH protocol

Motivated by the original LEACH and other improvement protocols [3, 18, 19], we proposed an optimization clustering method in LEACH protocol using modified k-means++ algorithm to overcome the energy consumption problems in WSN. We named this protocol as modified k-means++ for LEACH or MKPP-LEACH.

There are two main phases in MKPP-LEACH who still refer to the basic LEACH protocol. First phase is the clustering phase, when all the SNs should be clustered according to the MKPP-LEACH algorithm. This phase called as set-up phase, which will determine the CHs and CMs. The second phase is data transmission, wherein every CM will send data to the CH consecutively based on the TDMA schedule set by CH. Afterwards, CHs dispatch the data to the BS after their aggregation process is complete.

As mentioned before, k-means++ algorithm still used randomly selection for the first centroid. It might bring imprecise clusters if there is an inappropriate centroid selection. For this reason, we provoke an additional parameter for the first centroid selection: the furthest distance between SN and BS. Distance indirectly affects in energy consumption. The farther away, the more energy it takes. Hence, we put on the furthest node from BS as the initial centroid (C_1).

Hereinafter, the algorithm will calculate the distance from every node in the network to the C_1 and specify the next C_i for another cluster using Eq. (3). Concordance with basic LEACH protocol, every node will join to the nearest C_i. When the temporary cluster is created, the algorithm will check periodically whether there is a cluster centroid change. Eq. (4) is used to re-determine the cluster centre until there are no more cluster centre changes. All of this process will continually repeat until the cluster establishment is over.

Along with LEACH does not take into account the residual energy of the nodes during CH selection in set-up phase [26], we also proposed an energy-efficient scheme into our protocol by adjoining the residual energy of every nodes in the CH selection process. The nominated CH will be chosen in reference to the highest residual energy within the clusters, which is enumerated by Eq. (9).

$$\varepsilon_r = E_i - T_e \quad (9)$$

where ε_r indicates the node's residual energy for each round, E_i presents the energy of the *i* node in round, and T_e means for the total energy consumption of node in round.

As a conclusion, we optimize the set-up phase for LEACH protocol by generating an energy-saving method through pre-eminence clustering algorithm and CH selection, while on the whole steady-state phase still carry out by LEACH protocol. The detail procedure of MKPP-LEACH is described in Fig. 4.

5. Results & discussion

5.1 Experimental setup

In order to evaluate performance and lifetime of the proposed algorithm compared to basic LEACH [33], KM-LEACH [3, 18, 19] and MKPP-LEACH, we used MatLab simulation platform. We set up a simulation scenario with the parameters presented in the Table 1, which are used for most studies [3, 6, 11-12, 21, 23].

The other parameters were similar to those in literature, except for BS location and round periods,

MKPP-LEACH Algorithm for WSN

N : number of nodes
r_{max} : maximum round of set-up phase
k : cluster numbers
(p,q) : BS coordinate
m : network field (in meters)
d_i : the distance between SN to the BS
CC_k : the centroid of k cluster
C_1 : The furthest nodes from BS
d_j : the distance between SN to C_1
i : node in the network
ε_i : node residual energy
i_c : non-CH nodes
ε_0 : the initial energy for every node
ε_{TX} : energy dissipation for transmitter node
ε_{RX} : energy dissipation for receiver node

Inputs: N, r_{max}, k, C, i, ε
Outputs: a set of k clusters and data transmission

Initialization
1. Set N and r_{max}
2. Set k clusters and *(p,q)* coordinate
3. Initiate N nodes in $m * m$ unit

Set-up Phase
1. Calculate d_i using Eq. (3).
2. Determine the k cluster centroid (CC_k):
   Set the furthest node as C_1
   Calculate d_j using Eq. (3)
   For $d_j > d_{j+1}$
   Set d_j as CC_k
3. Allocate nodes to the nearest CC_k
4. For all nodes in network repeat below step.
   If there is a value change of CC_k,
   Then recalculate CC_{k-i} using Eq. (4)
   Else $i = CC_k$.
5. CH selection process:
   Calculate node's residual energy using Eq. (9)
   for $\varepsilon_i > \varepsilon_{i+1}$,
   i = CH

Steady-State Phase
1. **Calculate** ε_0 for every node before the data transmission is starts using Eq. (5)
2. For all nodes in the cluster repeat below step
   For *i=1:n*
   If ε_0 >= 0,
   send data to their CH
   calculate ε_{TX} residual energy using Eq. (5)
   else state this node as **dead node**
3. For all CH in the network repeat below step
   Send TDMA slots CH to the i_c
   Wait until all CM sent their data
   Calculate ε_{RX} residual energy using Eq. (6)
   If ε_{RX} >= 0,
   aggregate data from CM
   send data to BS
   else state this node as **dead node**

Figure. 4 Pseudocode of MKPP-LEACH Algorithm

Table 1. Simulation parameters

Parameters	Value
Number of nodes	100
Network field	100 x 100 m^2
Round periods	400
Number of clusters	8
Location of the BS	(50,50) m
CH selection probability	10%
Initial energy	0,5 J
Data length	4000 bits
Radio electronics energy	50nJ/bit
Multipath channel parameter	13nJ/bit/m^4
Free-space channel parameter	10nJ/bit/m^2
Data aggregation energy	50nJ/bit

depending on the purpose of each analysis. For instance, it was set to (100, 100) m in [23], (50, 150) m in [6, 11], and (50, 175) in [12, 21]. Moreover, we made some assumptions about the sensor field as follows: All of SNs are lifelong constructed and deployment (they are motionless nodes). Hereafter, they had the same initial energy.

5.2 Performance evaluation

Intending to evidence the feasibility of our proposed protocol, we delineate its performance evaluation into two main sections: the clustering formation and the network lifetime performance.

5.2.1. Nodes clustering formation

In this section, we will show the advantage of clustering algorithm incorporated for LEACH protocol in balancing the cluster formation. A total of 100 sensor nodes are randomly distributed for this study. The initial topology for WSN seen in Fig. 5, where the blue dot indicates the SNs while the red ones is the BS.

We have investigated the clustering formation from our proposed protocol and compared to the basic LEACH and KM-LEACH, as shown in Fig. 6. It is found that basic LEACH protocol produced unbalanced cluster, such as the red-o cluster only contains 4 nodes while the dark-blue ones contain 35 nodes (see Fig. 6(a)). The same issue was found in

Received: January 26, 2022. Revised: March 15, 2022.

Figure. 5 The initial topology for WSN

KM-LEACH (Fig. 6(b)), where the red-o cluster consist of 12 nodes while the dark-blue ones contain 10 nodes. But this imbalanced clusters were successfully ameliorated through our proposed protocol. We recognized that MKPP-LEACH constructs more balanced cluster, 12 nodes for both clusters (see Fig. 6(c)).

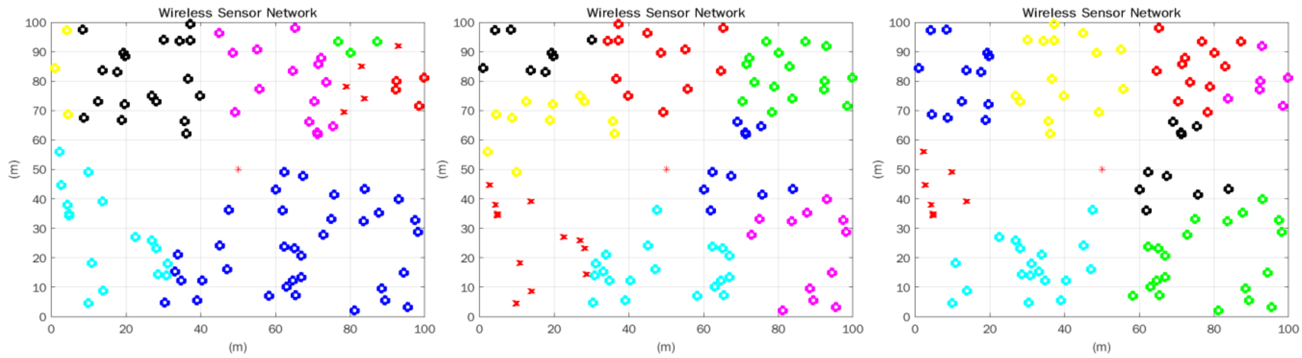
5.2.2. Network lifetime

The basic essential behind cluster optimization is to reach extensive network lifetime for SNs, considering this is the massive challenges in building WSN technologies. For this reason, we investigated the network lifetime for each protocol and figure out their performance in energy consumptions.

Fig. 7 represents total energy consumption during network lifetime. It is clear that our proposed routing scheme consumes less energy than LEACH and KM-LEACH protocol. This can be achieved through some reasons: (1) optimizing the initial cluster centres and classify the nodes in more balanced clusters (2) selecting CH with residual energy as consideration, which can prevent the CHs from dying too early and support a high number of communication rounds.

In fact, the most energy was expended when the nodes start to transmit their data in steady-state phase (either when CM send data to CH or CH to BS). This process may lead energy of the node to be exhausted and causes the node to die. Therefore, it is important to make sure that nodes will distribute evenly. An even number of nodes in a cluster had a positive impact in prolonging the CH's energy. Thereby reducing the node death risk due to lack of energy while data transmission is running.

As can be seen in Fig. 8(a) that the nodes are started to die when the rounds achieve more than 150 rounds for LEACH and 200 rounds for KM-LEACH. Nevertheless, MKPP-LEACH establishes node



(a) (b) (c)

Figure. 6 Clustering formation results: (a) LEACH (b) KM-LEACH (c) MKPP-LEACH

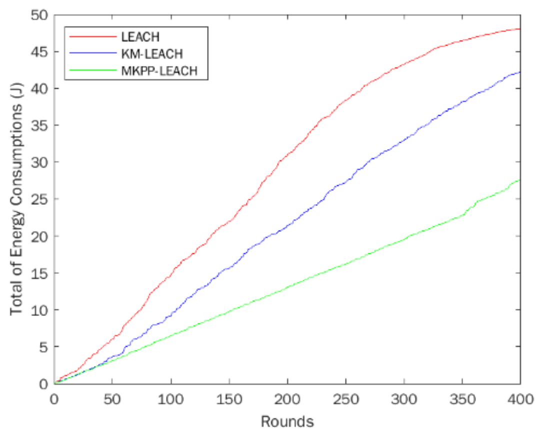
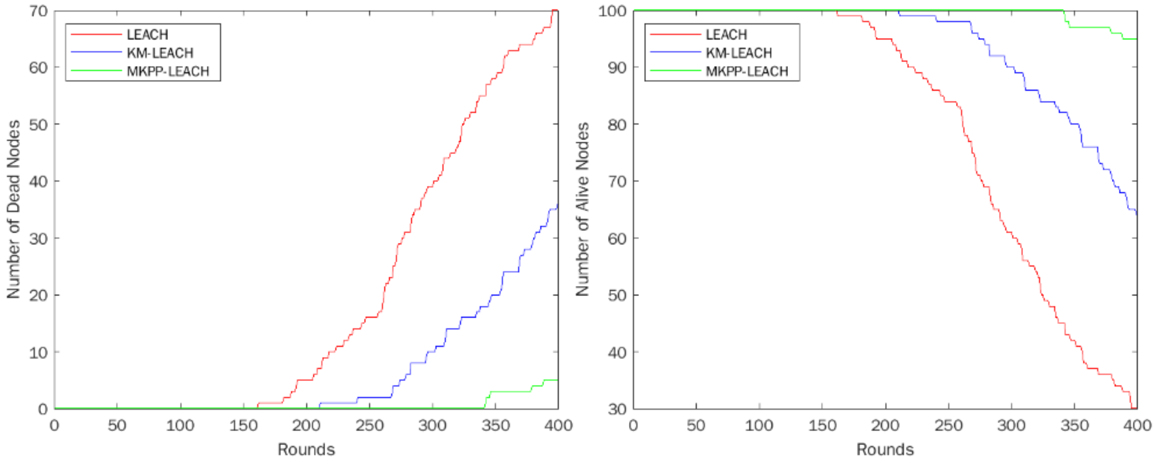


Figure. 7 Total energy consumptions (in Joule)

Received: January 26, 2022. Revised: March 15, 2022.



(a) (b)

Figure. 8 Network lifetime performance: (a) number of dead nodes (b) number of alive nodes

longer until more than 300 rounds. Obviously, MKPP-LEACH can keep most the nodes alive in the network (see Fig. 8(b)), which means it can balance the energy consumptions more effectively than LEACH or KM-LEACH. It is worthy to mention that MKPP-LEACH outperforms better performance in extending the network lifetime compared to LEACH and KM-LEACH.

6. Conclusions

In this paper, a clustering optimization based on modified k-means++ algorithm is proposed to accomplish the energy consumption problems in WSNs. The disadvantages of LEACH which is caused by unbalanced clusters and random CH selection are avoided. Furthermore, integrating additional clustering algorithm such as k-means not enough have significant impacts. A modified k-means++ algorithm named MKPP-LEACH is presented to figure out that drawbacks. The clustering and CH selection process are improved. The distance and residual energy of every node are considered as additional parameter in selecting the cluster centre and CH for each cluster. Using this idea, the total energy consumption and alive nodes are improved effectively. Our proposed protocol builds more balanced clusters, resulting in less total energy and longer network lifetime (up to 42,8 % compared to basic LEACH and 57,2 % for KM-LEACH).

Future works

However, there are some defects for MKPP-LEACH, such as larger memory space of nodes and longer process time for set-up phase compared to the basic LEACH. Another point for future works is adding the distance threshold as parameter in order to optimize the cluster head selection and balance the cluster formation respectively.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Fransiska Sisilia Mukti, Putu Mas Anggita Putra; methodology, Fransiska Sisilia Mukti, Putu Mas Anggita Putra; software, Putu Mas Anggita Putra; validation, Putu Mas Anggita Putra, Allin Junikhah; formal analysis, Fransiska Sisilia Mukti, Aryuanto Soetedjo and Awan Uji Krismanto; writing—original draft preparation, Fransiska Sisilia Mukti and Allin Junikhah; writing—review and editing, Aryuanto Soetedjo and Awan Uji Krismanto; visualization, Putu Mas Anggita Putra; project administration, Putu Mas Anggita Putra and Allin Junikhah.

Acknowledgments

This work was supported by the Ministry of Research and Technology/National Research and Innovation Agency, Indonesia under Inter-University Cooperation Research Grant No. 0167/SP2H/LT/DRPM/2021, 033/AMD-SP2H/LT-MULTI-PDPK/LL7/2021, 0167/B.1/LP2M/ITB-ASIA/IV/2021.

References

[1] L. Chan, K. G. Chavez, H. Rudolph, and A. Hourani, "Hierarchical Routing Protocols for Wireless Sensor Network: A Compressive Survey", *Wireless Networks*, Vol. 26, No. 5, pp. 3291–3314, 2020.

[2] R. E. Mohamed, A. I. Saleh, M. Abdelrazzak, and A. S. Samra, "Survey on Wireless Sensor

Network Applications and Energy Efficient Routing Protocols", *Wireless Personal Communications*, Vol. 101, No. 2, pp. 1019–1055, 2018.

[3] P. Saheb and Dr. K. Sharma, "Improved LEACH Protocol Based on K-Means Clustering Algorithm for Wireless Sensor Network", *International Journal of Electronics & Communication Technology*, Vol. 8, No. 4, pp. 28–32, 2017.

[4] S. Umbreen, D. Shehzad, N. Shafi, B. Khan, and U. Habib, "An Energy-Efficient Mobility-Based Cluster Head Selection for Lifetime Enhancement of Wireless Sensor Networks", *IEEE Access*, Vol. 8, pp. 207779–207793, 2020.

[5] I. Daanoune, B. Abdennaceur, and A. Ballouk, "A Comprehensive Survey on LEACH-based Clustering Routing Protocols in Wireless Sensor Networks", *Ad Hoc Networks*, Vol. 114, No. 1, pp. 1-21, 2021.

[6] J. G. Lee, S. Chim, and H. H. Park, "Energy-Efficient Cluster-Head Selection for Wireless Sensor Networks Using Sampling-Based Spider Monkey Optimization", *Sensors*, Vol. 19, No. 23, pp. 1-18, 2019.

[7] P. Singla and A. Munjal, "Topology Control Algorithms for Wireless Sensor Networks: A Review", *Wireless Personal Communications*, Vol. 113, No. 4, pp. 2363–2385, 2020.

[8] J. John and P. Rodrigues, "A Survey of Energy-Aware Cluster Head Selection Techniques in Wireless Sensor network", *Evolutionary Intelligence*, Vol. 11, No. 1, 2019.

[9] K. Guleria and A. K. Verma, "Comprehensive Review for Energy Efficient Hierarchical Routing Protocols on Wireless Sensor Networks", *Wireless Networks*, Vol. 25, No. 3, pp. 1159–1183, 2019.

[10] P. N. Prabha and A. Ali, "Energy Efficient Threshold Based Cluster Head Selection and Optimized Routing in LEACH", In: *Proc. of International Conference on Intelligent Data Communication Technologies and Internet of Things*, India, pp. 1475-1485, 2018.

[11] Y. Zhang, M. Liu, and Q. Liu, "An Energy-Balanced Clustering Protocol based on An Improved CFSFDP Algorithm for Wireless Sensor Networks", *Sensors (Switzerland)*, Vol. 18, No. 3, pp. 1-18, 2018.

[12] M. Elshrkawey, S. M. Elsherif, and M. E. Wahed, "An Enhancement Approach for Reducing the Energy Consumption in Wireless Sensor Networks", *Journal of King Saud University - Computer and Information Sciences*, Vol. 30, No. 2, pp. 259–267, 2018.

[13] Z. Y. Qiong and L. Y. Rui, "A Routing Protocol for Wireless Sensor Networks using K-Means and Dijkstra Algorithm", *International Journal of Advanced Media and Communication*, Vol. 6, No. 2–4, pp. 109–121, 2016.

[14] A. Shahraki, A. Taherkordi, Ø. Haugen, and F. Eliassen, "Clustering Objectives in Wireless Sensor Networks: A Survey and Research Direction Analysis", *Computer Networks*, Vol. 180, No. 3, 2020.

[15] T. N. Viet, T. P. T. Quynh, and H. D. Thi, "Enhancement of Leach based on K-Means Algorithm and Stochastic Optimization", *International Journal of Advanced Engineering Research.*, Vol. 16, No. 9, pp. 406–410, 2021.

[16] S. K. Singh, P. Kumar, and J. P. Singh, "A Survey on Successors of LEACH Protocol", *IEEE Access*, Vol. 5, No. 2, pp. 4298–4328, 2017.

[17] S. Y. Park, J. Y. Lee, and D. Lee, "Enhanced KOCED Routing Protocol with K-means Algorithm", *Computers, Materials and Continua*, Vol. 67, No. 3, pp. 4019–4037, 2021.

[18] M. Bidaki, R. Ghaemi, and S. R. K. Tabbakh, "Towards Energy Efficient k-MEANS based Clustering Scheme for Wireless Sensor Networks", *International Journal of Grid and Distributed Computing*, Vol. 9, No. 7, pp. 265–276, 2016.

[19] J. Y. Lee and D. Lee, "K-means Clustering-Based WSN Protocol for Energy Efficiency Improvement", *International Journal of Electrical and Computer Engineering*, Vol. 11, No. 3, pp. 2371–2377, 2021.

[20] B. Zhu, E. Bedeer, H. H. Nguyen, R. Barton, and J. Henry, "Improved Soft-k-Means Clustering Algorithm for Balancing Energy Consumption in Wireless Sensor Networks", *IEEE Internet of Things Journal*, Vol. 8, No. 6, pp. 4868–4881, 2021.

[21] S. Periyasamy, S. Khara, and S. Thangavelu, "Balanced Cluster Head Selection Based on Modified k -Means in a Distributed Wireless Sensor Network", *International Journal of Distributed Sensor Networks*, Vol. 2016, 2016.

[22] A. Ray and D. De, "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.

[23] B. M. Benmahdi and M. Lehsaini, "A Conditional Distributed K-means Cluster-based Routing Scheme for Wireless Sensor Networks", In: *Proc. of International Conference on*

Artifficial Intelligence and Information Technology, Singapore, pp. 121–126, 2019.

[24] R. Gantassi, B. B. Gouissem, and J. B. Othman, “Combination of MDC and K-Means to improve LEACH”, *IEEE International Conference on Communications*, Surabaya, Indonesia, pp. 101-109, 2021.

[25] N. Srikanth and M. S. G. Prasad, “Efficient Clustering Protocol Using Fuzzy K-Means and Midpoint Algorithm For Lifetime Improvement in WSNs”, *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 4, pp. 61–71, 2018, doi: 10.22266/ijies2018.0831.07.

[26] F. Zhao, Y. Xu, and R. Li, “Improved LEACH routing communication protocol for a wireless sensor network”, *International Journal of Distributed Sensor Networks*, Vol. 2012, 2012.

[27] H. Pambhar, K. Aghera, and N. Tada, “Energy efficient cluster head selection in energy-LEACH”, In: *Proc. of Advances in Intelligent Systems and Computing*, Singapore, pp. 197–204, 2017.

[28] N. D. Tan, L. Han, N. D. Viet, and M. Jo, “An Improved LEACH Routing Protocol for Energy-Efficiency of Wireless Sensor Networks”, *Smart Computing Review*, Vol. 2, No. 5, pp. 360-369, 2012.

[29] Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding”, In: *Proc of the Annual ACM-SIAM Symposium on Discrete Algorithms*, Stanford, pp. 1027–1035, 2007.

[30] S. Kumar, *Understanding K-Means, K-Means++ and, K-Medoids Clustering Algorithms*, Towards Data Science, New York, 2022.

[31] Geeks for Geeks, *K-means++ Algorithm*, Geeks, New York, 2022.

[32] O. Ogundile and A. Alfa, “A Survey on an Energy-Efficient and Energy-Balanced Routing Protocol for Wireless Sensor Networks”, *Sensors*, Vol. 17, No. 5, pp. 1084, 2017.

[33] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-Efficient Communication protocol for Wireless Microsensor Network”, In: *Proc of the 33rd Hawaii International Conference on System Sciences,* Maui, HI, USA, pp. 1-10, 2000.