INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 99-116, February 2018.

Dynamic Multi-hop Routing Protocol Based on Fuzzy-Firefly Algorithm for Data Similarity Aware Node Clustering in WSNs

M. Misbahuddin, A.A. Putri Ratna, R.F. Sari

Misbahuddin Misbahuddin*
Department of Electrical Engineering, Faculty of Engineering,
Universitas Indonesia, Depok, 16424, Indonesia
*Corresponding author: misbahuddin@ui.ac.id

Anak Agung Putri Ratna
Department of Electrical Engineering, Faculty of Engineering,
Universitas Indonesia, Depok, 16424, Indonesia
ratna@eng.ui.ac.id

Riri Fitri Sari
Department of Electrical Engineering, Faculty of Engineering,
Universitas Indonesia, Depok, 16424, Indonesia
riri@ui.ac.id

Abstract: In multi-hop routing, cluster heads close to the base station functionaries as intermediate nodes for father cluster heads to relay the data packet from regular nodes to base station. The cluster heads that act as relays will experience energy depletion quicker that causes hot spot problem. This paper proposes a dynamic multi-hop routing algorithm named Data Similarity Aware for Dynamic Multi-hop Routing Protocol (DSA-DMRP) to improve the network lifetime, and satisfy the requirement of multi-hop routing protocol for the dynamic node clustering that consider the data similarity of adjacent nodes. The DSA-DMRP uses fuzzy aggregation technique to measure their data similarity degree in order to partition the network into unequal size clusters. In this mechanism, each node can recognize and note its similar neighbor nodes. Next, K-hop Clustering Algorithm (KHOPCA) that is modified by adding a priority factor that considers residual energy and distance to the base station is used to select cluster heads and create the best routes for intra-cluster and inter-cluster transmission. The DSA-DMRP was compared against the KHOPCA to justify the performance. Simulation results show that, the DSA DMRP can improve the network lifetime longer than the KHOPCA and can satisfy the requirement of the dynamic multi-hop routing protocol.
Keywords: clustering, data similarity, multi-hop routing, fuzzy system, firefly algorithm, Wireless Sensor Networks (WSNs).

1 Introduction

Wireless Sensor Networks (WSNs) rapidly grow in various applications for many domains. Besides, WSNs is also an integral part of the Internet of Things (IoT) that can share data for improving the human capability in monitoring a local environmental condition and process automation. It consists of a set of sensor nodes that are deployed in an appropriate environment in the ad hoc model to observer and interact with the physical world or biological system remotely. Therefore, WSNs should be able to adapt dynamically with the charged environment. Recently, WSNs play important role in various necessaries of human such as flood monitoring [14], weather monitoring, earthquake detection [22], tracking [19], volcanic eruption, military necessity [10], healthcare observation [3] agriculture automation [26], and manufacturing automation [17].

Each node is composed of sensor, low ability processor, limited capacity storage and power supply, and transceiver. Therefore, the efficient usage of energy that is supplied by battery is important issue in order to prolong the network lifetime. Topologically, the clustering techniques have been commonly used to improve the network performance such as prolonging the network lifetime, enhancing the network scalability, increasing the bandwidth efficiency, and increasing fault tolerance [1]. Clustering technique divides the nodes on a network into many logical or physical groups termed as clusters. Each cluster is composed of a node selected as Cluster Head (CH) and many regular nodes called cluster members. Each regular node senses data of the environmental condition and forwards to its CH. Meanwhile, the CH functions to sense data, aggregate data, and relay them to the other CH or Base Station (BS).

Clustering techniques consist of two fashions, equal sized clustering and unequal sized clustering. In equal sized clustering, all clusters have the same size number of cluster members. The CHs closer to BS have an additional function, not only sensing data, aggregating data, and sending the aggregated data to BS but also forwarding data from the other CHs to BS. These CHs have a heavier load than the CHs farther from BS, so that they consume more energy and deplete energy more quickly than the other CHs. Thus, the network connectivity is disrupted in relaying data to BS. This event is termed as a hot spot problem.

To overcome the hot spot issue in the network, the topology of unequal sized clustering can be used to organize the load balancing among the CHs [?]. Architecture of the unequal sized clustering is to reduce the clusters size closer to BS and increase the clusters size as the distance between CH and BS. In our work, load of clusters can not be arranged through such way because the cluster size is determined according to the clustering technique based on the data similarity referred spatial and temporal correlation. Therefore, such clustering technique requires a specific routing protocol to increase the energy efficiency in transmitting the sensed data by the regular nodes to BS via either a CH with a single-hop or some CH with multi-hop. Furthermore, this technique is also a dynamically changed clustering in each round. The topology of the network changes in each round because each cluster is established based on the data similarity of the adjacent nodes.

Because such clustering technique generates unequal sized clusters, the selection of some nodes as CHs is a crucial problem for improving the energy saving in order to prolong the network lifetime. This paper proposed a dynamic multi-hop routing protocol designed specially for a data similarity aware node clustering that is a topology of the unequal sized cluster to improve the network lifetime, and satisfy the requirement of multi-hop routing protocol for the dynamic node clustering that consider the data similarity of adjacent nodes.

Generally, this protocol runs in two main steps. The first step is the dynamic node clustering based on the data similarity using fuzzy aggregation technique. The second step is a routing algorithm using the modified K-HOP Clustering Algorithm (KHOPCA) [6] by adding a priority factor that is obtained by a hybrid approach of fuzzy system and firefly algorithm. There are two variables that are considered to obtain the priority factor, i.e., the residual energy and distances between CHs and BS.

The remainder of this paper is organized as follows: Section 2 presents literature review related with the routing protocols in wireless sensor networks. Section 3 describes our approach used for dynamic multi-hop routing protocol. Section 4 presents the simulation results to show the performance evaluation. Finally, Section 5 concludes this paper.

2 Routing protocol in WSNs

Routing is the best path to transmit a data packet from a source node to a destination node. The clustering-based routing in WSNs, there are two types of path, i.e. data traffic within a cluster termed as intra-cluster, and data traffic between clusters called inter-cluster. In the intra-cluster, each regular node senses a local environmental condition and transmits it to corresponding CH. Meanwhile, the CH senses, receives, and aggregates data. Then it transmits the aggregated data packet to either BS directly or via intermediate CHs.

One of highly important issues in many literatures of the clustering-based routing techniques is the use of more energy efficient methods in order to prolong the network lifetime. There are two main steps that needs the appropriate technique in order to achieve better network performance in term of lifetime. Both steps are the clustering technique and the CHs election method. The clustering techniques are classified into two major categories, i.e. unequal sized clustering and equal sized clustering techniques. Similarly, the CHs election approaches are categorized into three major groups, i.e. preset, random, and attribute based method [1]. In preset approach, all nodes that are selected as CHs are adjusted before they were deployed in the environment. In random methods, the CHs are selected among of the nodes randomly in the field. On the other hand, attribute-based approaches select the CHs among of the nodes based on some of their characteristics, such as the residual energy and distance to the BS.

There are several equal clustering based routing protocols that select CHs randomly. Among other is the Low Energy Adaptive Clustering Hierarchy (LEACH) [12] that is a hierarchical clustering-based routing protocol which has been used widely as a benchmark. There are many LEACH-based routing protocols that have been proposed to improve the energy efficiency in order to prolong the network lifetime such as LEACH-Centralized (LEACH-C) [13], LEACH-based Energy (LEACH-E) [15], and LEACH with Distance-based Threshold (LEACH-DT) [16]. The disadvantage of the LEACH-based routing protocol is that CHs are close to the BS that consumes more energy than the CHs farther from BS. Consequently, the CHs near the BS died earlier. This causes a disruption of the network connectivity termed as a hot spot problem.

Single-hop routing protocol can overcome the hot spot problem because it does not require the intermediate BS to relay the data packet to BS. However, this approach has a limitation of transmission coverage, so that the scalability of the network cannot be achieved. In order to overcome the hot spot problem on the network, some unequal sized clustering approaches using single-hop routing have been proposed. However, these approaches are a waste of energy and the transmission coverage are limited [18]; [4]; [7]. Therefore, several multi-hop routing protocols using unequal sized clustering technique [2]; [21]; [24] have been proposed to overcome the hot spot problem, improve the network scalability and optimize the energy-saving in order to prolong the network lifetime.

In fact, some of specific applications in WSNs not only satisfy the three purposes, but also require a clustering technique based the data similarity readings of adjacent nodes to obtain the data pattern of the observed environment to make decision or prediction. Therefore, to fulfill the requirements of the applications, this paper proposes a new dynamic multi-hop routing protocol using the unequal size clustering technique based on the data similarity. This protocol proposes an incorporation the K-Hop Clustering Algorithm (KHOPCA) rules [6] and fuzzy system-firefly algorithm [25]. Our proposed routing protocol is called Data Similarity Aware for Dynamic Multi-hop Routing Protocol (DSA-DMRP). The DSA-DMRP starts to establish the node clustering based on data similarity among all nodes in the network using the Fuzzy Aggregation Technique as a dynamic unequal sized clustering mechanism. Finally, the CHs election and establishment of the best route path uses the KHOPCA rules and a priority factor in the CHs election. The priority factor considers the residual energy and distance to the BS using integration of Fuzzy System (FS) and Firefly Algorithm (FA) to improve the network lifetime.

3 The proposed model of routing protocol

3.1 Network model

Data similarity aware node clustering in WSNs is an unequal sized clustering-based WSNs. A node will merge or leave its cluster depend on its data similarity degree to those of neighbor nodes. Therefore, perhaps there are some nodes that are an intersection in some clusters. Due to an unequal sized clustering, there are some clusters that have member nodes fewer than those of other clusters. Moreover, there are some individual nodes which does not belong to any cluster. The characteristic of such a clustering-based model requires a properly specific routing algorithm. Hence, a multi-hop clustering-based routing protocol was developed to overcome the problem. All sensor nodes that are randomly deployed on the network are stationary. They are homogeneous in their capabilities of sensing, processing, and communicating. The BS is assumed as a stationary site, and it has no energy constraint. Not all nodes can communicate directly with other nodes. Only nodes that can communicate directly to the sink will be considered to become the cluster heads. The node clustering and the data gathering are run into rounds. In each round, there are four steps, (i) all nodes sensed the local data, (ii) the nodes divided themselves into several clusters, (iii) the nodes created a multi-hop routing, (iv) the member nodes sent their data to the corresponding CHs, and the CHs aggregates the data in order to forward them to the BS.

3.2 Energy consumption model

Energy consumption in WSNs is an urgent issue to be considered in order to increase the longer network lifetime. The energy consumption models have been developed in several literatures. Figure 1 shows the consumption model of radio energy.

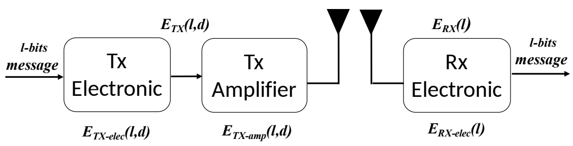

Figure 1: Transmission energy dissipation model [24]

The dissipated energy to transmit *l-bits* message with distance d and receive *l-bits* can be computed respectively using the following equations [7].

$$\begin{aligned} E_{TX}(l,d) &= E_{TX-elec}(l) + E_{TX-amp}(l,d) \\ &= l(E_{elec} + E_{amp}d^p) \\ &= \begin{cases} l(E_{elec} + \varepsilon_{fs}d^2, & if \quad d \le d_0 \\ l(E_{elec} + \varepsilon_{mp}d^4, & if \quad d > d_0 \end{cases} \end{aligned} \tag{1}$$

$$E_{RX}(l) = E_{RX-elec}(l) = l \times E_{elec} \tag{2}$$

where E_{elec} is the consumed energy in either transmitter or receiver circuit, is for amplifier radio model, and p is the path loss. The path loss is adjusted to $p = 2$ and the free space model is used by assigning $E_{amp} = \varepsilon_{fs}$ if the distance d is less than or equal to the threshold distance. On the contrary, the path loss is assigned $p = 4$ and the multi-path fading model is employed by setting $E_{amp} = \varepsilon_{mp}$ if distance d more than the distance threshold d_0. Besides, the sensing data by sensor also consumes amount of energy significantly as follows [24].

$$E_{Sensing}(l) = l \times E_{sens} \tag{3}$$

The clustering model based on data similarity has a high correlation between the data read by each node in any cluster. Therefore, CH can use data aggregation ways to bundle all highly related data into a single length-fixed packet. The consumed energy to aggregate n packet of *l-bits* by the cluster head is calculated by [24]:

$$E_{aggre}(l) = n \times l \times E_{DA} \tag{4}$$

In the network, there are K clusters in which each cluster contains of s_i $(i = 1,\ 2,\ ...,\ K)$ regular node. In each round, a node senses *l-bit* packet data and transmits it to the corresponding cluster head once a frame. Thus, the consumed energy by a regular node in the intra-cluster communication between the regular node i and its cluster head j is calculated by [24]:

$$E_{reg-j}(i,l,d) = E_{Sensing}(l) + E_{TX}(l,d) \tag{5}$$

where *d(i, j)* is the distance between regular node i and CH j. On the other hand, the consumed energy by the CH in both intra-cluster and inter-cluster communication consist of five activities i.e. CH senses data, CH receives data packets, CH aggregate data packets, CH receives and transmits data packets from the other CH, and CH transmits the aggregated packets and the received packets of other CH to the BS [24]. Thus, the consumed energy by a CH can be obtained using the following equation [24].

$$\begin{aligned} E_{CH}(i,l,d) = l(&E_{Sensing}(l) + E_{elec}s_i + E_{DA} \\ &(s_i+1) + E_{elec}relay(j) + (relay(j)+1) \\ &(E_{elec} + E_{amp}d(i,NH))_p) \end{aligned} \tag{6}$$

where *d(i, NH)* is the distance between CH j and next hop (NH) in which NH may be CH or base station. Moreover, *relay(j)* is the number of forwarded packets from other CHs.

3.3 Dynamic node clustering based on data similarity

In WSNs, there are several applications that require a data similarity based node clustering approach. Such approach is highly related to two important issues, i.e. the spatial and temporal correlation. In the spatial correlation, the data that is sensed by adjacent nodes tend to have a high data similarity degree. Meanwhile, in the temporal correlation, the data that is sensed by each node consecutively at an observed location tend to have a high correlation of its data readings.

The fuzzy aggregation technique in equation 7 [9] can measure the data similarity degree considering the spatial correlation of two data a and b that is sensed by two adjacent nodes.

$$Sim(a,b) = exp\frac{-||a-b||^2}{2 \times \eta^2}, \quad Sim(a,b) = \in [0,1] \tag{7}$$

where η is the constant Gaussian Kernel by setting $\eta = 1.74$. The data similarity degree of 1 is a highest level and 0 is a lowest level. To establish the unbalanced clusters based on the data similarity, the DSA-DMRP use two main algorithms to identify the neighbor nodes that satisfy the data similarity degree. Both algorithms are embedded in each node in order to broadcast and receive the beacon message as shown in Algorithm 1 and 2 [20] :

1. In each node, two simple data structures are required: (i) the *SpatNeighbor* is a vector data structure to store the information of a spatially closed neighbor nodes; and (ii) *simNeighbor* is a vector data structure to store the information of the similar neighbor nodes. The information contains all of the three local variables, i.e., address (*add*), current data readings (*cdR*), and the number of similar neighbors (*nsN*).

2. Algorithm 1: Beacon message, the **Broadcasting** (line 1-3) is a main function to broadcast a beacon message periodically. This function consists of three sub functions: (i) the Send sub function (line 1) transmits the information about the identifier address (*add*), the current data reading (*cdR*), and the number of similar neighbors (*nsN*). (ii) the Delay sub function prevents the transmissions simultaneously. (iii) the TimeExpire sub function limits the broadcasting time of the nodes.

3. Algorithm 2: The **Receiving** main function is employed by the node to receive a beacon message. The message is used by any node receiver to identify whether the transmitter node includes as a similar neighbor node.

4. The beacon message of *add*, *cdR*, and *nsN* is utilized to identify the similar neighbor candidates (line 1-2). The data similarity degree is measured by the fuzzy aggregation technique (line 7). If its similarity degree is more than or equal to *SiDegree* and also the similar neighbor candidate is not existed within the data structure, it is added into the data structure as one of the members of the similar neighbor (line 10). The number of similar neighbor (*nsN*) is incremented (line 11). In contrast, if the similar neighbor candidate existed within the data structure, it is removed from the data structure (line 13), and the number of similar neighbor (*nsN*) is decremented (line 14).

Algorithm 1 Beacon Message

Broadcasting()

```
1: Send(add, cdR, nsN)
2: Delay(interval + rand())
3: TimeExpire()
```

3.4 Cluster head election and multi-hop routing

DSA-DMRP forms routes that are started through selecting several cluster heads. Establishing routes and selecting CHs utilize K-Hop Clustering Algorithm (KHOPCA) [6] that consists of four rules. However, the rules proposed by the KHOPCA do not consider how to prolong the network lifetime and overcome the hot spot problem. To address both problems, our DSA-DMRP proposed modified KHOPCA's rules by adding a Priority Factor (PF) to select the prospective common nodes as the CHs. The PF is calculated via an incorporation between fuzzy model and Firefly Algorithm (FA) that consider two input variables, i.e. residual energy and distance to the base station.

The KHOPCA constructs the network routes using a set rules that were inspired by Game of Life [8]. The KHOPCA is implemented over the networks that have been clustered. A route from a regular node to the base station is established based on a set of rules that define transition of previous state to current state of a node depending on the previous state of its similar neighbors. The state of a node is represented by a weight $w \in [wMin, wMax]$. Minimum distance to a cluster head is defined as the minimum weight *wMin*, whilst the maximum weight *wMax* is the

Algorithm 2 Receiving data for node clustering

Receiving*(add,cdR,nsN)*

spatNeighbor
simNeighbor
simNeighbor $\leftarrow$ CreateSimNeighbor(add, cdR, nsN)
sigma $\leftarrow$ 1.74
n $\leftarrow$ cdR
m $\leftarrow$ DataReading()
s $\leftarrow$ exp(- || n-m || ^{2}/(2 * sigma^2)) Eq. 7
if (s $\geq$ SiDegree) **then**
    if (!IsExistSimNeighbor(simNeighbor) **then**
        AddSimilarNeighbor(simNeighbor)
        IncreNumSimilarNeighbor()
    else if (**then**IsExistSimNeighbor(simNeighbor)
        RemoveSimNeighbor(simNeighbor)
        DecrNumSimNeighbor()
    end if
end if

maximum distance to a cluster head. The KHOPCA has four rules to establish the network route as follows [6]:

1. If node i with wight w_i has a neighbor node that has the highest weight w_n where $w_n > w_i$, ∀n ∈ *LN(i)* of its all neighbor nodes, and *LN(i)* is the list of i similar neighbors, the node i changes its weight to $w_i = w_n$ - 1.

2. If node i has no a similar neighbor with a higher weight, where $w_i \neq wMax$, $w_i \geq w_n$ and $\forall w_n \in W(LS(i))$, the node i decreases its weight to $w_i = w_i$ - 1

3. If weight of the node i is $w_i = wMin$ and also $w_i = w_n$ where $w_n \in W(LS(i))$, the node i adjusts its weight to $w_i \leftarrow wMax$ and states itself as CH. In this case, none of its similar neighbors has a higher weight than $wMin$.

4. If weight of node i is $w_i = wMax$ and weight of one of its neighbor nodes has also weight $w_n = wMax$, where $\exists\ w_n \in W(LS(i))$, the node n decreases its weight to $w_n = w_n$ - 1. In this case, there are two CHs in the same cluster.

Although those four rules are simple, they can construct all nodes in the network to create a multi-hop routing. The first rule aims to form a top-down hierarchical structure through adjusting its weight with a difference-one of the highest weight node existing in the list of similar neighbors. The second rule intends to avoid the less weighted nodes that are most likely to quit from a cluster. Thus, the higher weighted nodes that the CH attract at the surrounding nodes will merge into its cluster in order to a fragmented cluster. The third rule describes that a node declares itself as a CH if all similar neighbors have a minimum weight. This situation shows that the isolated nodes are chosen as CH on the minimum weight level. The fourth rule overcomes a situation where there are two CHs in the same cluster. Therefore, one of them must survive as a CH, while other node must be a follower node of the CH.

The weakness of the rules is that it has not considered how to prolong the network lifetime. Therefore, the third rule determines a node to be a CH. The fourth rule defined itself as a CH.

Both rules are modified by adding a priority factor that considers the residual energy and the distance to the BS as follows.

3. If weight of the node i is $w_i = wMin$, $w_i = w_n$ where $w_n \in W(LS(i))$, and also its PH is highest in the same cluster, the node i adjusts its weight to $w_i \leftarrow wMax$ and states itself as CH. In this case, none of its similar neighbors has a higher weight than $wMin$.

4. If weight of node i is $w_i = wMax$, and weight of one of its neighbor nodes has also weight $w_n = wMax$, where $\exists\ w_n \in W(LS(i))$, as well as its PH is highest in the same cluster, the node n decreases its weight to $w_n = w_n$ - 1. In this case, there are two CHs in the same cluster.

Priority factor of cluster head selection

One of criteria to select prospective regular node as CH is the priority factor as shown in third and fourth rule of modified KHOPCA's rules. The Priority Factor (PF) is calculated using a incorporation between Fuzzy Logic and Firefly Algorithm (FA). The PF is obtained through a procedure that consists of four steps: normalization, fuzzification, inference engine, and defuzzification as shown in Figure 2. To obtain the proper selection of fuzzy rule in inference process, firefly algorithm is used to optimize the the Tsukamoto fuzzy rule base table. As input variable of the fuzzy logic, our proposed DSA-DMRP considers two variables, i.e. the residual energy $E_r(n)$ and the distance to the BS $d_{BS}(n)$.

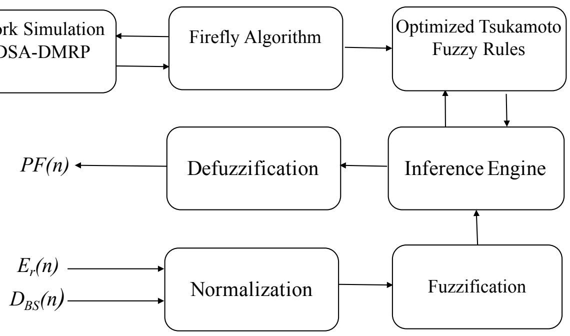

Figure 2: Procedure of priority factor to select cluster head

Before fuzzification that maps crisp input to membership degree via membership function, the first step normalizes both input variables E_r(n) and D_{BS} (n) into range of [0, 1]. This step is performed to avoid the difference of range value in each cluster. Normalization of both input variables use a general formula as follows:

$$d^n(n) = \frac{d(n) - min(d)}{max(d) - min(d)} \tag{8}$$

where $d^n(n)$ is normalized value and $d(n)$ is real values of input variable d for node n. The input variable d can be applied for both input variables E_r and D_{BS}. The maximum real value of d is defined as $max(d)$ and the minimum real value of d is represented as $min(d)$. In the second step, the fuzzifier maps normalized crisp values d to the membership functions to convert to be the linguistic fuzzy values. In this study, our proposed fuzzy model uses five membership functions that are symbolized with Very Low *(VLow)*, *Low*, *Medium*, and *High* for both input variables as shown in equation 9 up to 13 respectively.

$$y_{VLow} = \begin{cases} 1 & if \quad x \leq 0.1 \\ \frac{0.3-x}{0.3-0.1} & if \quad 0.1 \geq x \leq 0.3 \\ 0 & if \quad x \geq 0.3 \end{cases} \tag{9}$$

$$y_{Low} = \begin{cases} 0 & if \quad x \leq 0.1 \; or \; x \geq 0.5 \\ \frac{x-0.3}{0.3-0.1} & if \quad 0.1 \geq x \leq 0.3 \\ \frac{0.5-x}{0.5-0.3} & if \quad 0.3 \geq x \leq 0.5 \end{cases} \tag{10}$$

$$y_{Madium} = \begin{cases} 0 & if \quad x \leq 0.3 \; or \; x \geq 0.7 \\ \frac{x-0.5}{0.5-0.3} & if \quad 0.3 \geq x \leq 0.5 \\ \frac{0.7-x}{0.7-0.5} & if \quad 0.5 \geq x \leq 0.7 \end{cases} \tag{11}$$

$$y_{High} = \begin{cases} 0 & if \quad x \leq 0.5 \; or \; x \geq 0.9 \\ \frac{x-0.7}{0.7-0.5} & if \quad 0.5 \geq x \leq 0.7 \\ \frac{0.9-x}{0.9-0.7} & if \quad 0.7 \geq x \leq 0.9 \end{cases} \tag{12}$$

$$y_{High} = \begin{cases} 0 & if \quad x \leq 0.7 \\ \frac{0.9-x}{0.9-0.7} & if \quad 0.7 \geq x \leq 0.9 \\ 1 & if \quad x \geq 0.9 \end{cases} \tag{13}$$

Likewise, the defuzzifier use seven membership functions i.e. Very Small (*VSmall*), *Small*, Rather Small (*RSmall*), *Medium*, Rather Large (*RLarge*), *Large*, and Very Large (*VLarge*) as shown in equation 14 up to 20 respectively to obtain the fuzzy output.

$$PF_{VSmall} = \begin{cases} 1 & if x \leq 0.05 \\ \frac{0.2-x}{0.2-0.05} & if \quad 0.05 \geq x \leq 0.2 \\ 0 & if \quad x \geq 0.2 \end{cases} \tag{14}$$

$$PF_{Small} = \begin{cases} 0 & if \quad x \leq 0.05 \; or \; x \geq 0.35 \\ \frac{x-0.2}{0.2-0.05} & if \quad 0.05 \geq x \leq 0.2 \\ \frac{0.35-x}{0.35-0.2} & if \quad 0.2 \geq x \leq 0.35 \end{cases} \tag{15}$$

$$PF_{RSmall} = \begin{cases} 0 & if \quad x \leq 0.05 \; or \; x \geq 0.35 \\ \frac{x-0.35}{0.35-0.2} & if \quad 0.2 \geq x \leq 0.35 \\ \frac{0.5-x}{0.5-0.35} & if \quad 0.35 \geq x \leq 0.5 \end{cases} \tag{16}$$

$$PF_{Madium} = \begin{cases} 0 & if \quad x \leq 0.35 \; or \; x \geq 0.65 \\ \frac{x-0.5}{0.5-0.35} & if \quad 0.35 \geq x \leq 0.5 \\ \frac{0.65-x}{0.65-0.5} & if \quad 0.5 \geq x \leq 0.65 \end{cases} \tag{17}$$

$$PF_{RLarge} = \begin{cases} 0 & if \quad x \leq 0.5 \; or \; x \geq 0.8 \\ \frac{x-0.65}{0.65-0.5} & if \quad 0.5 \geq x \leq 0.65 \\ \frac{0.8-x}{0.8-0.65} & if \quad 0.65 \geq x \leq 0.8 \end{cases} \tag{18}$$

$$PF_{Large} = \begin{cases} 0 & if \quad x \leq 0.65 \; or \; x \geq 0.95 \\ \frac{x-0.8}{0.8-0.65} & if \quad 0.65 \geq x \leq 0.8 \\ \frac{0.95-x}{0.95-0.8} & if \quad 0.8 \geq x \leq 0.95 \end{cases} \tag{19}$$

$$PF_{VLarge} = \begin{cases} 0 & if \quad x \leq 0.8 \\ \frac{0.95-x}{0.95-0.8} & if \quad 0.8 \geq x \leq 0.95 \\ 1 & if \quad x \geq 0.95 \end{cases} \tag{20}$$

In third step, the inference engine performs a fuzzy reasoning against the crisp input in the fuzzy rule base table containing n rules. In our study, we use Tsukamoto Fuzzy system [23] with two inputs and an output. The typical rules of Tsukamoto uses AND-based fuzzy rule base table that are represented as IF-THEN as shown in Equation (21) as follows:

$$\textbf{IF} \quad in_1 = A \ \textbf{AND} \ in_2 = B \ \textbf{THEN} \ out = C \tag{21}$$

where A and B are the membership degree of the corresponding input membership functions and C is min(A, B).

In final step, all outputs of the fired rules are aggregated and converted to be a single-crisp output value. To obtain the single-scrip output value as value of the priority factor, our fuzzy model uses the average-weighted Tsukamoto defuzzification model [23]. The Priority Factor *PF(n)* can be formulated in Equation (22) as follows:

$$PF(n) = \frac{\Sigma_{i=1}^{25} \mu_{i \times C_i}}{\Sigma_{i=1}^{25} \mu_i} \tag{22}$$

where μ_i is min(μ_{Er}, $\mu_{D_{BS}}$) to corresponding membership functions within the fuzzy rule i. Also, C_i is the output of corresponding membership function in rule i.

Optimization of AND-based fuzzy rule via firefly algorithm

The Inference engine of fuzzy system has usually many rules. The selection of fuzzy rule requires a proper method to obtain the best performance of the fuzzy system. In our fuzzy model, there are two input variabels in which each input variable has five membership functions. These mean that the number of rules is 5 x 5 = 25. Because the output fuzzy has seven member functions, the number of output alternatives of the 25 rules is 7^{25}. The output alternatives of 7^{25} become an NP-hard problem to find the best solution in turning of the fuzzy rule base table. The NP-hard problem can be addressed a fuzzy system that uses Firefly Algorithm (FA) to optimize the Tsukamoto fuzzy rule base table in order to prolong the network lifetime based data similarity aware node clustering.

FA is a population-based swarm intelligent search algorithm [11]. Each individual firefly in population has a role as a candidate solution in the search space. Each firefly moves toward a new position. The new position represents a better candidate solution. Finally, they find the best solution. The movement is represented by their attractiveness. The attractiveness is proportional to the intensity of the emitted light by adjacent fireflies. The better solution is usually measured by the fitness value.

Let x_i be the ith firefly in the population, where i = 1, 2, ..., N and N is the population size. The attractiveness β with the Euclidian distance r_{ij} between two adjacent fireflies x_i and x_j can be computed using the Equation (23) as follows [27]:

$$\beta(r_{ij}) = \beta_{min} + (\beta_0 - \beta_{min})e^{-\gamma r_{ij}^2} \tag{23}$$

$$r_{ij} = ||x_i - x_j|| = \sqrt{\sum_{k=1}^{D} (r_{ik} - r_{jk})^2} \tag{24}$$

where D is the problem dimension with $k = 1, 2, \ldots, D$. The parameter β_0 indicates the fireflies' attractiveness at the distance $r = 0$, and γ is the light absorption coefficient. β_{min} is the minimum value of β as shown in Equation 23. The attractiveness of β is limited in the range of $[\beta_{min}, \beta_0]$. Each firefly X_i is compared with all other fireflies x_j , where $j = 1, 2, \ldots, N$ and . If x_j is brighter than x_i, x_i will be attracted to and move toward x_j. The movement of the firefly x_i that is attracted toward the firefly x_j can be calculated by [27]:

$$x_{ik}(t+1) = x_{ik}(t) + \beta^{-\gamma r_{ij}^2} + (x_{ik} - x_{jk} + \alpha(t) s_i \varepsilon_i \tag{25}$$

$$\alpha(t+1) = \alpha(t)(1/9000)^{\frac{1}{t}} \tag{26}$$

where ε_i is a uniformly distributed random value in the range of [-0.5, 0.5] and parameter α is the dynamically updated step factor using Equation (26) and s_i is the length of scale of each designed variable.

In Algorithm 3 [28], the FA starts to optimize the AND-based fuzzy rules through generating randomly the population of N fireflies. Since there are 25 controllable parameters in fuzzy rules as mentioned previously, the length of feasible solution is a string of 25. Each value of feasible solution contains number 1 to 7 representing seven output member functions.

Algorithm 3 FA to optimize the AND-based fuzzy rules

Input: Population: N; Dimension D; Iterator time: T Output: Global best firefly's brightness x(t)

```
t ← 1 (initialization)
initialize all fireflies brightness x_i^k
while t ≤ T do
    Update the parameter α according to Eq. 26
    for i=1:N do
        for j=1:N do
            if j ≠ i then
                Compute the attractiveness of β
                according to Eq. 23
                if f(x_j(t) < f(x_i(t)) then
                    Move x_i(t) toward x_j(t)
                    according to Eq. 25
                    f(x_i(t)) ← Evaluate Fitness of Firefly
                    according to Eq. 27
                    t ← t + 1
                end if
            end if
        end for
    end for
end while
```

Before conducting the iteration process, the fitness value of each feasible solution is computed using $f(X_i)$ fitness function. The optimal solution will be obtained when the iteration reaches the maximum iteration time M. In each iteration, the parameter is updated firstly using Equation (26). Next, the attractiveness β between two fireflies x_i and x_j is calculated according to Equation (23), where $j \neq i$. The movement of the x_i firefly towards the x_j firefly using Equation (25) is processed if the fitness value of x_j firefly is better than that of x_i firefly.

To evaluate each feasible solution, the FA requires a fitness function *f(x_i)*. The feasible solutions are taken from the corresponding Tsukamoto fuzzy rules to be extracted and simulated in the network using Network Simulation of DSA-DMRP. The best solution will be obatined if all fitness within same approaching value. The fitness is computed using three parameters, i.e. the number of rounds when the first node dies (FND), the number of rounds in which half of nodes are dead (HND), and the number of rounds until the last node dies (LND). The network lifetime is measured using the three parameters. The Equation (27) is the fitness function and its constrains, which are used to maximized the network lifetime formulated as follows [29]:

Maximize:

$$Fitness = w_1 FND + w_2 HND + w_3 LND \tag{27}$$

Address to

$$0 \leq w_j \leq 1, \qquad \sum_{j=1}^{3} w_j = 1 \tag{28}$$

where w_j with j = (1, 2, 3) are the weights to determine the important objective of parameters of *FND*, *HND*, and *LND*.

4 Performance evaluation

This study is a experimental research using a a NS-3 network simulator version 3.25. The performance of our proposed DSA-DMRP was evaluated using term of the network lifetime of *FND*, *HND*, and *LND*, as well as the round history of alive nodes in various data similarity degrees of the cluster. Furthermore, the DSA-DMRP was compared agianst the KHOPCA with exactly the same scenario.

4.1 Simulation scenario

The experimental data used to simulate the both protocols utilized the humidity readings gathered by the Intel Berkeley Research Lab [5]. The data was collected from 54 sensor nodes deployed in a 640m x 480m sized network. Before the simulation is executed, there are some the network parameters and setting parameters of firefly algorithm that need to be set as shown in Table 1 and Table 2 respectively.

Table 1: Simulation Parameters of Networks

Parameter	**Value**
Data similarity degree (siDegree)	0.7 to 0.9
Gaussian Kernel constant η	1.74
Initial energy	0.5 Joule
E_{elec}	50 nJ/bit
ε_{fs}	100 pJ/bitt/m^2
ε_{amp}	0.03 pJ/bit/m^2
Data packet size	4000 bit

Table 3: Comparison between KHOPCA and proposed DSA-DMRP in term of the network lifetime and in various data similarity degrees (siDegree)

	KHOPCA			DSA-DMRP		
SiDegree	FND	HND	LND	FND	HND	LND
0.7	206	1393	1421	543	1489	1575
0.8	134	1368	1391	316	1420	1477
0.9	240	1380	1452	316	1368	1470

Table 2: Setting parameters of firefly algorithm

Parameter	**Value**
Maximum iterations time (T)	100
The number firefly in population (N)	30
β_0	1
β_{min}	0.2
ε_{amp}	0.005
Dimension of population (D)	25

4.2 Experimental results

Table 3 shows the results obtained in the comparison between KHOPCA and our proposed DSA-DMRP in term of *FND*, *HND*, and *LND*. In this experiment, there are three weights to set the fitness function i.e. w_1=0.8, w_2=0.2 and w_3=0. Moreover, the node clustering based on data similarity was established through three experiments with the data similarity degree (*SiDegree*) 0.7, 0.8, and 0.9.

The DSA-DMRP was compared against the KHOPCA to justify the performance. It shows that the DSA-DMRP can reach a longer network lifetime than the KHOPCA in all conditions due to an addition of the priority factor in the KHOPCA's rules. However, the difference between *FND* and *HND* is highly significant both in the KHOPCA and the DSA-DMRP because the unequal sized cluster is not designed specially to overcome the hot spot problem for CHs close to BS but actually, it is designed in other to satisfy the requirement of the multi-hop routing protocol for the dynamic node clustering based on the data similarity of their neighbors.

Figure 3 shows three round histories of alive nodes versus rounds for each routing protocol with *SiDegree* 0.7, 0.8 and 0.9 respectively. It clearly shows that the DSA-DMRP and the KHOPCA have an approximately same stability of alive nodes. Both protocols show that most of nodes died simultaneously in term of LND. However, the DSA-DMRP is longer in all terms FND, HND, and LND. Therefore, the DSA-DMRP can extend the network lifetime in a relatively significant manner. The stability of alive nodes in both protocols are caused by the stability of KHOPCA's rules in selecting the CHs. Meanwhile, the network lifetime in DSA-DMRP that longer than the KHOPCA is caused by adding the factor priority in selecting CHs.

In order to justify the stability of alive nodes in each protocol for three data similarity degrees, we compare them in variable the data similarity degree for each protocol. Figure 4 shows the round history of alive nodes for the KHOPCA and the DSA-DMRP. The KHOPCA clearly shows the data similarity degree of 0.7 and 0.8 are more stable than at of 0.9. Whereas, the DSA-DMRP indicates that the data similarity degree of 0.8 and 0.9 is more stable than those of 0.7.

However, Figure 4 shows an inverse phenomenon between both the alive node stability graphs. The phenomenon of the KHOPCA protocol shows the graphs of similarity degrees of 0.7 and 0.8 that are more stabile than the similarity degree of 0.9. On the other hand, in the phenomenon

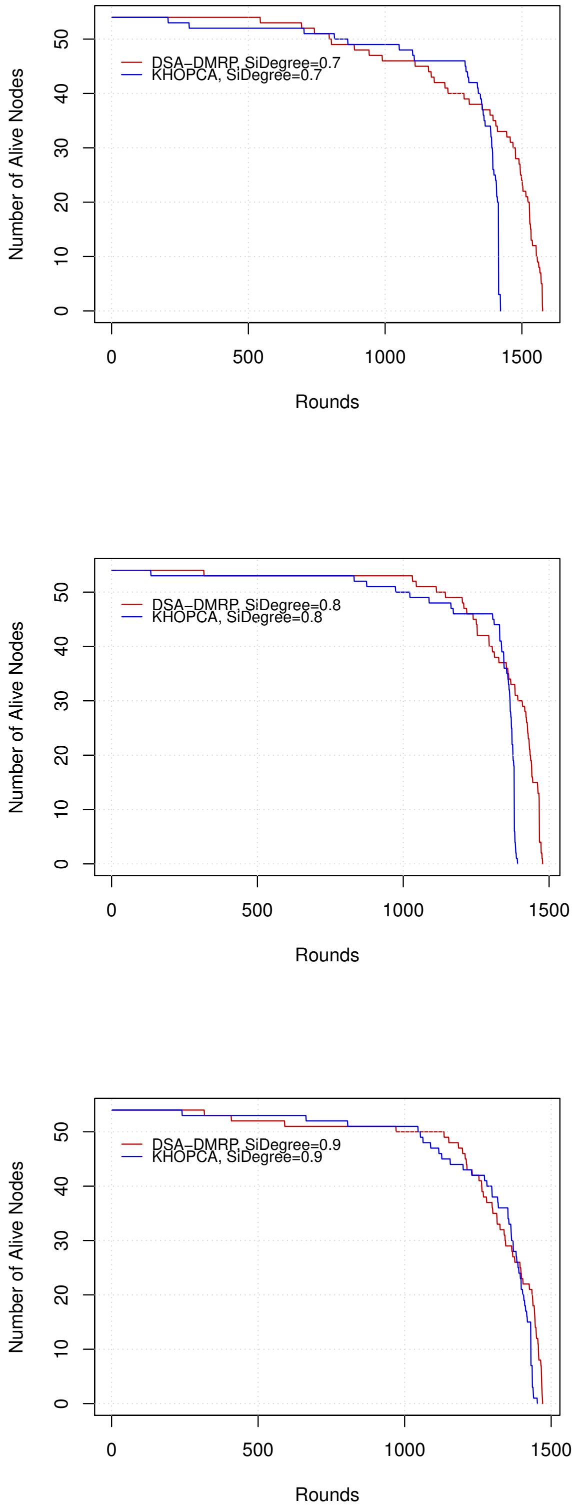

Figure 3: Relation between the number of alive nodes and rounds in $SiDegree = 0.7$, 0.8, and 0.9

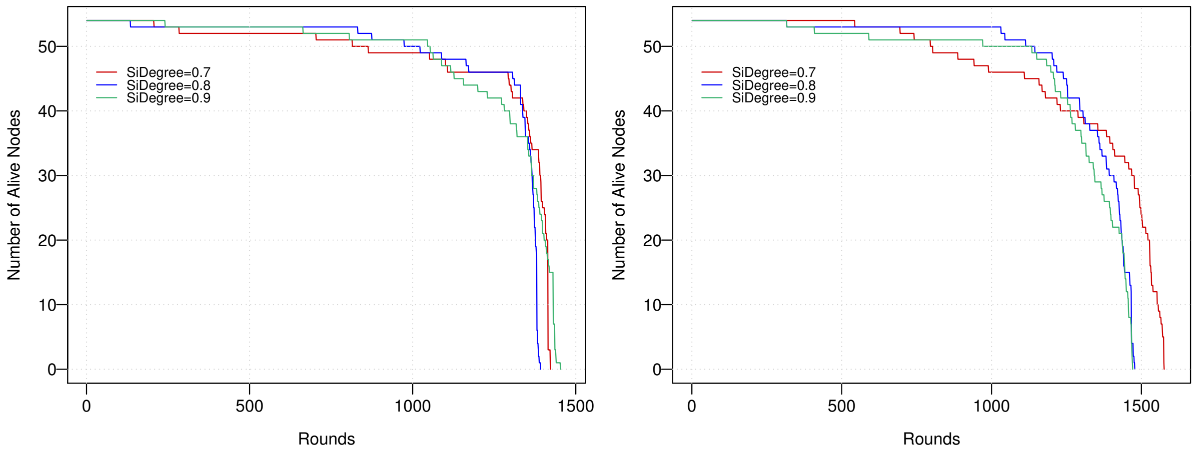

Figure 4: Comparation between the data similarity degree in the alive nodes vs the rounds for the KHOPCA and the DSA-DMRP

of DSA-DMRP protocol, the similarity degrees of 0.8 and 0.9 are more stable than the similarity degree of 0.7. Moreover, in the rounds before reaching 1000 of the KHOPCA protocol, the number of formed clusters in the similarity degrees of 0.7, 0.8, and 0.9 are almost same. Thus, more and more cluster heads in the similarity degree of 0.9, more and more the probabilities of the nodes will be dead because the cluster heads more consume the energy than those of the member nodes. On the contrary, the rounds before reaching 1000 of the DSA-DMRP, the number of formed clusters is almost same the number of formed clusters in the KHOPCA protocol. However, before reaching 1000 rounds, the DSA-DMRP has less the number of cluster heads in the similarity degrees of 0.8 and 0.9 than those of the similarity degree of 0.7. Finally, the problems in this case are caused by the elected cluster heads that are not only effected by the similarity degree but also affected by the residual energy and the distance to the base station.

5 Conclusions

Our proposed DSA-DMRP is a dynamic multi-hop routing protocol using unequal sized clustering approach. This protocol is based on the modified KHOPCA rules by adding a priority factor. The priority factor is a parameter for selecting the CHs in the network that consider the residual energy and distance to the BS. The fuzzy aggregation technique is used to measure the data similarity degree of adjacent nodes.

The DSA-DMRP was compared against the KHOPCA to justify the performance. The DSA-DMRP and the KHOPCA have an approximately same stability of alive nodes. However, The DSA-DMRP can reach a longer the network lifetime than the KHOPCA in all terms of FND, HND, and LND. Therefore, the DSA-DMRP can extend the network lifetime in a relatively significant manner and can satisfy the requirement of multi-hop routing protocol for dynamic node clustering based on the data similarity of their neighbors.

Acknowledgement

The authors gratefully acknowledge the Ministry of Research, Technology and Higher Education of Indonesia for supporting this work through DIKTI's research grant under grant No. 1173/UN2.R12/HKP.05.00/2016

Bibliography

[1] Afsar, M. M.; Tayarani, M. H. (2014); Clustering in sensor networks: A literature survey, *Journal of Network and Computer Applications*, 46, 198-226, 2014.

[2] Ahmed, N.; Kanhere, S.; Jha, S. (2010); Experimental Evaluation of Multi-hop Routing Protocols for Wireless Sensor Networks, *In Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 416–417, New York, NY, USA, 2010.

[3] Amin, R.; Islam, S. H.; Biswas, G. P.; Khan, M. K.; Kumar, N. (2016); A robust and anonymous patient monitoring system using wireless medical sensor networks, *Future Generation Computer Systems*, 2016.

[4] Bagci, H.; Yazici, A. (2013); An energy aware fuzzy approach to unequal clustering in wireless sensor networks, *Applied Soft Computing*, 13(4), 1741-1749, 2013.

[5] Bodik, P., Hong, W., Guestrin, C., Madden, S., Paskin, M., and Thibaux, R., Intel Lab Data. Retrieved from http://db.csail.mit.edu/labdata/labdata.html.

[6] Brust, M. R.; Frey, H.; Rothkugel, S. (2008); Dynamic Multi-hop Clustering for Mobile Hybrid Wireless Networks, *Proceedings of the 2Nd International Conference on Ubiquitous Information Management and Communication*, 130-135. New York, NY, USA: ACM, 2008

[7] Gajjar, S.; Talati, A.; Sarkar, M.; Dasgupta, K. (2015); FUCP: Fuzzy based unequal clustering protocol for wireless sensor networks, *39th National Systems Conference (NSC)*, 2015.

[8] Gardner, M. (1970); Mathematical Games: The Fantastic Combinations of Jhon Conway's New Solitaire Game "Life", City, 1970.

[9] Ghaddar, A.; Razafindralambo, T.; Simplot-Ryl, I.; Tawbi, S.; Hijazi, A. (2010); Algorithm for data similarity measurements to reduce data redundancy in wireless sensor networks, *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on*, 1-6, 2010.

[10] Gupta, R.; Sultania, K.; Singh, P.; Gupta, A. (2013); Security for wireless sensor networks in military operations, *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on*, 1-6, 2013.

[11] Haibin, D.; Qinan, L. (2015); New progresses in swarm intelligence-based computation, *Int. J. Bio-Inspired Computation*, 7(1), 26-35, 2015.

[12] Heinzelman, W. R.; Chandrakasan, A.; Balakrishnan, H. (2000); Energy-efficient communication protocol for wireless microsensor networks, *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 3005-3014, 2000.

[13] Heinzelman, W. B.; Chandrakasan, A. P.; Balakrishnan, H. (2002); An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications*, 2002.

[14] Horita, F. E. A.; Albuquerque, J. P. de; Degrossi, L. C.; Mendiondo, E. M.; Ueyama, J. (2015); Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks, *Computers and Geosciences*, 80, 2015.

[15] Jia, J.G.; He, Z.W.; Kuang, J.M;, Mu, Y.H. (2010); An Energy Consumption Balanced Clustering Algorithm for Wireless Sensor Network, *2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010.

[16] Kang, S. H.; Nguyen, T. (2012); Distance Based Thresholds for Cluster Head Selection in Wireless Sensor Networks, *IEEE Communications Letters*, 16(9), 1396-1399, 2012.

[17] Kollam, M.; Shree, S. R. B. S. (2011); Zigbee Wireless Sensor Network for better Interactive Industrial Automation, *Advanced Computing (ICoAC), 2011 Third International Conference on*, 304-308, 2011.

[18] Li, C.; Ye, M.; Chen, G.; Wu, J. (2005); An energy-efficient unequal clustering mechanism for wireless sensor networks, *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference*, 597-608, 2005.

[19] Lu, K.; Zhou, R.; Li, H. (2016); Event-triggered cooperative target tracking in wireless sensor networks, *Chinese Journal of Aeronautics*, 29(5), 1326-1334, 2016.

[20] Misbahuddin, M.; Sari, R. F. (2016); Data Similarity Based Dynamic Node Clustering Using Bio-Inspired Algorithm for Self-Organized Wireless Sensor Networks, In *P. Novais and S. Konomi (Eds.), Intelligent Environments*, 318-327, London, UK: IOS Press, 2016.

[21] Purkait, R.; Tripathi, S. (2015); Fuzzy based unequal energy aware clustering with multi-hop routing in wireless sensor network, *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, 2015.

[22] Rahman, M.; Rahman, S.; Mansoor, S.; Deep, V.; Aashkaar, M. (2016); Implementation of ICT and Wireless Sensor Networks for Earthquake Alert and Disaster Management in Earthquake Prone Areas, *Procedia Computer Science*, 85, 92-99, 2016.

[23] Ross, T. J. (2010); *Fuzzy Logic with Engineering Applications*, Mexico, USA: Jhon Wiley & Sons, 2010.

[24] Sabor, N.; Abo-Zahhad, M.; Sasaki, S.; Ahmed, S. M. (2016); An Unequal Multi-hop Balanced Immune Clustering protocol for wireless sensor networks, *Applied Soft Computing*, 43, 372-389, 2016.

[25] Saleem, M.; Di Caro, G. A.; Farooq, M. (2011); Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions, *Information Sciences*, 181(20), 4597-4624, 2011.

[26] Tuan Dinh, L.; Dat Ho, T. (2015); Design and deploy a wireless sensor network for precision agriculture. In Information and Computer Science (NICS), *2015 2nd National Foundation for Science and Technology Development Conference on*, 294-299, 2015.

[27] Xin-She, Y. (2008); Cuckoo Search and Firefly Algorithm Xin-She Yang Editor Theory and Applications. (X.-S. Yang, Ed.). London, UK: Springer. http://doi.org/10.1007/978-3-319-02141-6, 2008.

[28] Wang, H.; Wang, W.; Zhou, X.; Sun, H.; Zhao, J.; Yu, X.; Cui, Z.(2017); Firefly algorithm with neighborhood attraction, *Information Sciences*, 382-383, 374-387, 2017.

[29] Zahedi, Z.M.; Akbari, R.; Shokouhifar, M.; Safaei, F.; Jalali, A. (2016); Swarm intelligence based fuzzy routing protocol for clustered wireless sensor networks, *Expert Systems with Applications*, 55, 313-328, 2016.