A Chain-Based Wireless Sensor Network Model Using the Douglas-Peucker Algorithm in the IoT Environment

Se-Jung LIM

Abstract: WSNs which are the major component in the IoT mainly use interconnected intelligent wireless sensors. These wireless sensors sense monitor and gather data from their surroundings and then deliver them to users or access connected IoT devices remotely. One of the main issues in WSNs is that sensor nodes are generally powered by batteries, but because of the rugged environments, it is difficult to add energy. The other one may cause an unbalanced energy consumption among sensor nodes due to the uneven distribution of sensors. For these reasons, the death of nodes by the energy exhausting and the performance of the network may rapidly decrease. Hence, an efficient algorithm study for prolonging the network lifetime of WSNs is one of important challenges. In this paper, a chain-based wireless sensor network model is proposed to improve network performance with balanced energy consumption via the solution of the long-distance communication problem. The proposed algorithm is consisted of three phases: Segmentation, Chain Formation, and Data Collection. In segmentation phase, an optimal distance tolerance is determined, and then the network field is divided into small sub-regions according to its value. The chain formation is started from the sub-region far away from the sink, and then extended, and sensed data are collected along a chain and transmitted to a sink. Simulations have been performed to compare with PEGASIS and Enhanced PEGASIS using an OMNET++ simulator. The simulation results from this study showed that the proposed algorithm prolonged the network lifetime via the achievement of the balanced energy consumption compared to PEGASIS and Enhanced PEGASIS. The proposed algorithm can be used in any applications to improve network performance of WSNs.

Keywords: BCFA; Chain Formation; Data Collection; Douglas-Peucker line-simplification algorithm; Internet of Things (IoT); PEGASIS; Wireless Sensor Networks (WSNs)

1 INTRODUCTION

The Internet of Things (IoT) is enabled by the latest developments in radio frequency identification (RFID), smart sensors, communication technologies, and Internet protocols. The basic premise is to have smart sensors collaborate directly without human involvement to deliver a new class of applications [1]. Another foundational technology for the IoT is the wireless sensor networks (WSNs), which mainly use interconnected intelligent sensors for sensing and monitoring. It can construct many applications (disaster relief applications, environment control and biodiversity mapping, intelligent buildings, facility management, machine surveillance and preventive maintenance, medicine and health care, logistics, telematics, and so on) [2, 3]. WSNs are used for collecting data from their surroundings for delivering them to users and accessing connected IoT devices remotely. They comprise the extensive number of small sensor nodes that can detect, compute, and communicate with other devices [4]. Sensor nodes are generally powered by batteries, and it is difficult to add energy to the sensor nodes due to the rugged environments where they operate. Therefore, they must at least operate for a given mission time or as long as possible. Besides, the uneven distribution of sensors may cause an unbalanced energy consumption among sensor nodes. Also, it can lead to rapid energy exhaustion of some nodes [5]. Because of the node death by energy exhausting, the performance of the network such as connectivity, coverage, lifetime, etc. may rapidly decrease [6, 7]. For this reason, the study of energy-efficient network model is one of the important challenges in WSNs. The chain-based routing algorithm, one of the important network models, can get the following strong points [8, 9]. Firstly, it saves more energy than cluster-based topologies do. Secondly, its energy distribution is even. Thirdly, it offers a longer lifetime for WSNs with low power consumption. Finally, it reduces the overhead coming from dynamic cluster formation. Many researchers have made improved algorithms based on the chain to prolong the network lifetime of WSNs in [10-17]. Among them, Power-Efficient Gathering in Sensor Information Systems (PEGASIS) is the most well-known routing algorithm based on a greedy chain formation approach in WSNs. However, in PEGASIS, some nodes consume much more energy to transmit data to their neighbour so that their energy will deplete very quickly, resulting in a decrease of the network lifetime.

In this paper, a chain-based wireless sensor network model (WSNM) is studied to avoid the unbalanced energy consumption by the long-distance communication of sensor nodes and prolong the network lifetime in WSNs, and the main contributions are as follows: 1) The Douglas-Peucker line-simplification algorithm is adopted to divide the network field into small sub-regions. 2) The chain formation in the sub-region far away from the sink based on each sub-region is started and extended to the adjacent sub-region. 3) Sensor nodes communicate with their immediate neighbours along the chain to collect data. The performance of the proposed algorithm is proved by comparison with the existing chain-based routing algorithms from the simulation experiment using OMNET++ [21, 22].

The remainder of this paper is organized as follows. Related work is presented in Section 2. In Section 3, the proposed algorithm is described in detail. The simulation results are discussed and compared in Section 4. Finally, this paper concludes in Section 5.

2 RELATED WORKS

2.1 PEGASIS and Related Research

PEGASIS [10] is one of the most well known routing protocols in WSNs. The main idea is that each node receive from and transmit to near neighbours along a chain and take turns becomes the leader for transmitting data to a BS. In PEGASIS, sensor nodes are randomly deployed in the network field and organized into a chain using the greedy algorithm as shown in Fig. 1 before the first round. If there is a sensor node already included in the chain, it is
not able to revisit the sensor node. In each round, a leader node among sensor nodes gathers data along a chain from the end nodes using the token passing mechanism. Each sensor node fuses its data and data received from its neighbour in the chain and transmits the fused data towards a leader node. The leader node sends the fused data to the BS as shown in Fig. 2.

COSEN [11] uses multiple chains. These chains contain multiple levels. In each level, the chain is formed using the greedy algorithm and a leader node is selected. A leader node of each lower-level collects information from its chain and sends the information towards a leader of the higher-level through the routing paths. The leader node of the higher-level transmits the collected information to the sink. In Enhanced PEGASIS [12], the authors formed the concentric clustering scheme based on multiple chains to gather data efficiently. Sensor nodes are assigned their level in the form of the concentric circle by the signal strength from the BS. In a chain of different levels, the head node is selected by turns. It informs its location information to the head node of the upper level and the lower-level and sends the gathered data to the head node of the lower-level. Finally, the head node of the lowest level transmits data to the BS. DS-PEGASIS [13] use one or two head nodes in the alternate concentric cluster according to levels. Levels with two head nodes collect data and send to the lower-level cluster consisting of one head node, and then the level consisting of one head node sends its data to the next level cluster with two head nodes. The data packets are transmitted to the BS along head nodes as diamond-shaped structures. In [14], the authors proposed diverse strategies of leader selection such as random, shuffle, and high-energy and 2-blocks and 4-blocks approach to maximize the lifetime of the nodes in the chain formed by the greedy algorithm. The random and shuffle strategies do not consider the available energy at the nodes, but the high-energy, 2-blocks and 4-blocks strategies consider. In [15], the authors proposed a way to address the data loss problem caused by the data transmission failure of a leader node. When the leader node for each round cannot transmit aggregated data to the BS, data loss will occur. One node among the leader node’s neighbours is selected based on the residual energy and transmits the aggregated data to the BS without loss of data. In BCBRP[16], the authors partition the network into equal size sub-areas and employ the chain structure using an algorithm similar to the minimum spanning tree algorithm. In each sub-area, nodes are connected in a chain, and chains are connected using bridge nodes. In EPEGASIS [17], since energy consumption rises rapidly with the increasing of the communication distance, multi-hop transmission was adopted to conserve energy. Optimal communication and optimal hop counts were used to decrease the times of data forwarding and achieve an optimal energy consumption. Also, EPEGASIS includes threshold value, communication distance, and mobile sink technology.

2.2 Douglas-Peucker Line-Simplification Algorithm

The Douglas-Peucker line-simplification algorithm [18, 19] is the most commonly used global line-simplification algorithm in cartography and Geographic Information Systems (GIS). It is recognized as the one that delivers the best perceptual representations of the original lines and is used extensively for both computer graphics and geographic information systems [20]. As shown in Fig. 3, if $Vs$ and $Ve$ are respectively the start and end vertices of the original entire line, an initial line segment $VsVe$ means the single-chain which joins two vertices. The perpendicular distance from all intermediate vertices to that initial line segment $VsVe$ is calculated. If the vertex has the maximum perpendicular distance from an initial line segment $VsVe$ and is larger than a specified tolerance, this vertex has served as a key of the simplification in the next step. The initial line segment $VsVe$ is calculated. If the vertex has the maximum perpendicular distance from an initial line segment $VsVe$ and is larger than a specified tolerance, this vertex has served as a key of the simplification in the next step. In other words, the initial line segment $VsVe$ is segmented into two. This process is repeated until the perpendicular distance for all vertices of the entire original line is smaller than the specified tolerance. The intermediate vertex within the specified tolerance is eliminated and then the vertices that have the maximum value of perpendicular distances is connected as results of simplification.
3 PROPOSED ALGORITHM

As the uneven distribution of sensors in one of the main issues in WSNs, which are the major component in the IoT, may cause the unbalanced energy consumption, this leads to the rapid energy exhaustion of sensor nodes.

The aim of this study is to prolong the network lifetime via the achievement of the balanced energy consumption among sensor nodes. The proposed algorithm will consider the distances between a sink and sensor nodes, then forming a chain by using the regional division repeatedly. The advantage is that once the chains are formed, there is no set-up overhead at the beginning of each round. In the proposed algorithm, a chain is created using the centralized chain formation algorithm. The chain formed by the sink will in general be better than those formed using the distributed algorithm. Sensor nodes send information about its current location to the sink before the first round. This information may be obtained by using the Global Positioning System (GPS) receiver that is activated. The sink determines optimal chains and broadcasts the chain information to the nodes. This broadcast message includes its ID of each node. For example, if there are N sensor nodes in the network field, sensor nodes are assigned to an integer in the interval (1, n) as the identifier names. Until the typical death of the node caused by energy exhaustion occurred, the chain will be held. In other words, when the sensor node communicates with other nodes by the long-distance for receiving and sending the data packet in the chain, the chain cannot be maintained by the death of nodes caused by energy exhaustion. The proposed algorithm constructs the network architecture based on the following methods: network field segmentation, chain formation, and data collection. When sensor nodes are non-uniformly distributed in the field of the network, then the network field is segmented into small regions based on basic concepts of the Douglas-Peucker line-simplification algorithm. Its function is as follows to segment the network field:

Starting node. This node is located the farthest from the sink and will be the first node in the chain.

Ending node. This node is the farthest sensor node from the starting node.

Segmentation line. It is a straight-line linked between the starting and ending nodes and segments the network field.

Optimal distance tolerance. This value is an optimal distance to segment both sides-minimal sensing areas based on the segmentation line.

Perpendicular distance. It is a distance which is a line perpendicular to sensor nodes from the segmentation line.

Perpendicular node. This node is a sensor node with the maximal distance among the perpendicular distances larger than the tolerance where $PD_4$ and $PD_5$ are larger than the tolerance. Because $PD_4$ was the maximal distance, $PD_4$ will be a perpendicular node.

Tentative line. It is a straight line drawn to the perpendicular node from starting and ending nodes. They will perform the role of the segmentation line to more segmentation in the network field.

Sensor nodes form a small chain in each region and then extend their chain. For forming a chain, sensor nodes within each small region form individual chains, and then they link the nearest perpendicular node to extend their chain. To collect data, a chain leader forwards data collected at sensor nodes to the sink.

3.1 Segmentation

The network field segmented into small regions using the basic concepts of the Douglas-Peucker line-simplification algorithm is introduced in the section 2. The segmentation of the network field can be made in the four-step procedure.

Step 1. The most important thing is to determine an optimal distance tolerance. Since the tolerance has a significant effect on a communication distance between sensor nodes, the optimal distance tolerance allocation based on the location of alive nodes is important.

Step 2. If there are starting and ending nodes, a straight-line, i.e., segmentation line, between two nodes is linked as shown in Fig. 4. This line segments the network field into two smaller regions.

Step 3. Intermediate nodes except for the starting and ending nodes calculate a perpendicular distance from the segmentation line. Sensor nodes with the perpendicular distance greater than the optimal distance tolerance will be candidate nodes of a perpendicular node, then one node

![Figure 4 Network field segmentation](image)
located at the long-distance between them is set up a perpendicular node. The perpendicular line, linking the perpendicular node and segmentation line, can segment the network field into smaller regions. The perpendicular line, linking the perpendicular node and segmentation line, can segment the network field into smaller regions. For example, the perpendicular line \( PL_4 \) can segment the network field into two regions (region\(_1\) and region\(_2\)) as follows:

\[
\text{region}_1 = \{\text{node}_1, \text{nodes}_5, \text{node}_7\} \\
\text{region}_2 = \{\text{node}_2, \text{node}_3, \text{node}_4, \text{node}_6, \text{node}_7\}
\]

**Step 4.** If the links between a perpendicular node and starting and ending nodes are connected, the two links become new segmentation lines. The network field will be divided into smaller, closer regions between sensor nodes by new segmentation lines. Every node except in perpendicular, starting, and ending nodes calculates a new perpendicular distance from new segmentation lines and then a new perpendicular node is selected based on the tolerance. Until the perpendicular node is no longer in existence, step 2, 3, and 4 are repeated.

### 3.2 Chain Formation

Fig. 5 describes the formation and extension flows of local chains. The formation process of the local chain starts from the starting node, and then the nearest node will be selected as the next node within a small region using the greedy algorithm. When a local chain is formed, the nearest perpendicular node to extend their local chain is selected as the next hop node and then starts all over again the chain formation process. The chain formation process continues until all sensor nodes are joined to the chain.

![Figure 5 Workflow of chain formation](image)

3.3 Data Collection

The discussion point in this section is how to collect data and transmit to a sink, and the same manners as in PEGASIS are used to collect data. One node among sensor nodes in the chain for each round takes turns becoming the chain leader for transmitting data to the sink. To elect a chain leader in the rth round, formula \( i \mod n \) (i: node ID, n: number of sensor nodes) is used. The token passing mechanism is used to collect data towards the chain leader from the starting and ending nodes along the chain. Since the token is very small, it does not affect the network lifetime. The sensor node fuses its neighbour's data with its data and then transmits to the other neighbour along the chain. The data fusion is performed at every sensor node except starting and ending nodes in the balanced chain. The chain leader commonly receives data from both neighbours and transmits the fused data to the sink. In the proposed algorithm, the network field is sectionalized into small areas, and a chain is formed based on these small regions. In the chain, each sensor node communicates with near neighbours to collect data also and all sensor nodes take turns becoming the chain leader. Therefore, the proposed algorithm can reduce the unbalance of energy consumption by the long-distance communication and prolong the network lifetime. The rotation of the chain leader that communicates directly with the sink is performed by the specific period of time to save energy of sensor nodes.

4 SIMULATION

The performance of the proposed model discussed in the previous section is evaluated. Simulations are driven by OMNET++ [21, 22] simulator and run on Windows 10. The network lifetime of WSNM to define the optimal distance tolerance is measured. WSNM is compared with PEGASIS and Enhanced PEGASIS via the following metrics.

**Long-distance communication.** This means the longest distance for receiving and sending the data packet in the chain.

**Network lifetime.** It is the total number of rounds that the sink receives the data from the chain leader until the first node dies.

**Average remaining energy.** It means average values for the remaining energy of alive nodes when the first node dies.

Assumptions and scenarios are the following:

- All proprietary rights relating to the article, other than copyright, such as patent rights.
- Homogeneous sensor nodes are randomly distributed in the network field.
- All sensor nodes are aware of their location (i.e. GPS signals).
- They are stationary after distribution and can send data to a sink and adjust their transmission range.
- They cannot be recharged, and energy is restricted.

In the network topology for the simulation as shown in Fig. 6, 100, 150, 200, and 250 sensor nodes are non-uniformly distributed within 100 m × 100 m network size. The radio model for transmitting and receiving a message that adopts the same radio model used in PEGASIS and Enhanced PEGASIS is defined in Tab. 1. In Tab. 1, \( E_{TX}(k) \) is energy consumption for transmitting a k-bit message.
over a distance, \(d\), and \(E_{RX}(k)\) is the formula to compute energy consumption for receiving a \(k\)-bit message. \(E_{elec}\) is the amount of energy consumed in the electronic circuit to transmit or receive the signal, and \(C_{amp}\) is the consumed energy in amplifiers. In the case of \(E_{TX}(k, d)\), the distance \(d\) assumes \(d^2\) energy loss caused by channel transmission. In Tab. 2, the parameters for the radio model in Tab. 1 are defined.

\[
E_{TX}(k, d) = E_{TX-elec}(k) + E_{TX-amp}(k, d)
\]

\[
E_{RX}(k) = E_{elec} \times k
\]

\[
E_{RX}(k) = E_{elec} \times k
\]

4.1 Simulation Results

4.1.1 Network Lifetime Analysis of WSNM

The scenario 1 to find out an optimal distance tolerance in WSNM is defined in Tab. 3. The network lifetime of WSNM for the different number of nodes according to the scenario 1 is measured. Fig. 7 shows the performance results for the network lifetime in WSNM.

In simulation results of the scenario 1, when the total number of sensor nodes and the optimal distance tolerance are respectively 250 and 30, the results indicate a more noticeable effect. Therefore, the optimal distance tolerance is set up 30 to compare with PEGASIS and Enhanced PEGASIS.

4.1.2 Long-Distance Communication Analysis

There is a close relation between energy consumption and communication distance, such as the parameters defined in Tab. 2. Fig. 8 indicates measurement results of the long-distance communication between sensor nodes under different algorithms based on the scenario 2 which is defined in Tab. 4.

In the case of WSNM (tolerance = 30), when the number of sensor nodes, \(N\), is 100, 150, 200, and 250, the results show around 82, 62, 51, and 37 meters, respectively. Enhanced PEGASIS shows overall similar results in
comparison with WSNM (tolerance = 30). Contrastively, the long-distance communication of PEGASIS increased approximately from 20% up to 50% compared to WSNM (tolerance = 30).

### Table 4 Scenario 2

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEGASIS,</td>
<td>Algorithms</td>
</tr>
<tr>
<td>Enhanced PEGASIS,</td>
<td>WSNM (tolerance =30)</td>
</tr>
<tr>
<td>(50, 300)</td>
<td>Sink’s location</td>
</tr>
</tbody>
</table>

4.1.3 Network Lifetime Analysis

When a sink is located at (50, 200), (50,300), and (50, 400), Fig. 9a, Fig. 9b, and Fig. 9c respectively indicate the performance of the network lifetime under different algorithms. The simulated results by the scenario 3 which is defined in Tab. 5, each algorithm shows similar results in every case, irrespective of the sink location. WSNM shows the increased network lifetime of approximately 2 to 5 times more efficient than PEGASIS and Enhanced PEGASIS and the first death of sensor nodes came later than other algorithms. These results mean that the proposed algorithm achieved the balanced energy consumption via the solution of the long-distance communication problem.

### Table 5 Scenario 3

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEGASIS,</td>
<td>Algorithms</td>
</tr>
<tr>
<td>Enhanced PEGASIS,</td>
<td>WSNM (tolerance =30)</td>
</tr>
<tr>
<td>(50, 200), (50, 300), (50, 400)</td>
<td>Sink’s location</td>
</tr>
</tbody>
</table>

4.1.4 Average Remaining Energy Analysis

Fig. 10a, Fig. 9b, and Fig. 9c illustrates an average remaining energy based on the scenario 3 in Tab. 5. In simulation results, when the death of the first node comes, PEGASIS and Enhanced PEGASIS used less energy than WSNM. It means the unbalanced energy consumption of sensor nodes and the rapid energy exhaustion of some node. Also, concerning the analysis of the previous network lifetime, the network lifetime of WSNM was approximately 2 to 5 times more efficient than PEGASIS and Enhanced PEGASIS. All of these results show that WSNM achieved more balanced energy consumption.
The proposed algorithm was demonstrated by comparing various simulation scenarios, the performance of the distance communications between sensor nodes. From balanced energy consumption via the prevention of long-chain. Consequently, the proposed algorithm achieved the can exchange data with near neighbour nodes along the formed based on the sub-region and extended, sensor nodes the network field into small sub-regions. Since a chain was Tehnički vjesnik 28, 6(2021), 1825-1832

5 CONCLUSION

WSNs are one of foundation technologies in IoT and mainly use the interconnected intelligent sensors for sensing and monitoring. In the uneven distributed WSNs, the rapid energy exhaustion caused by the unbalance energy consumption among sensor nodes is causative of the network performance decrement. In this paper, a chain-based wireless sensor network model (WSNM) was proposed to prolong the network lifetime of WSNs in the IoT environment. In the proposed algorithm, the Douglas-Peucker line-simplification algorithm was adopted to divide the network field into small sub-regions. Since a chain was formed based on the sub-region and extended, sensor nodes can exchange data with near neighbour nodes along the chain. Consequently, the proposed algorithm achieved the balanced energy consumption via the prevention of long-distance communications between sensor nodes. From various simulation scenarios, the performance of the proposed algorithm was demonstrated by comparing PEGASIS and Enhanced PEGASIS. The future work will be mainly centered on the evaluation and the optimization.

Acknowledgements

This study was supported by research fund from Honam University, 2020.

6 REFERENCES


Contact information:

Se-Jung LIM, Assistant Professor
Liberal Arts & Convergence Studies,
Honam University,
120, Honamdae-gil, Gwangsan-gu, Gwangju-si, 62399, Republic of Korea
E-mail: limsejung@honam.ac.kr