ABSTRACT
Designing a Wireless Sensor Networks (WSN) mostly was a great challenge. Shown in previous results, some design approaches lead to problems in its implementation. Deterministic methods face the NP-Hard complex problem. On the other side, heuristic methods sometimes produce a flawed result. With those situations, this research concern with exploring the possibility of a multi-objective optimization (MOO) method. As with the MOO method, some conflicted WSN aspects consider simultaneously. Started with the PSO algorithm, this developing method tries to find the best position of the WSN’s relays. Closed neighbor sensor nodes are then will be connected. It is combined with the graph to constructs the best communication link. These steps will be done in a certain number of iterations to enhance fault-tolerance ability. This MOO approached method was implemented to different WSN topologies, with several sensors placed in a simulation area. Used as controls are Steiner Point and Triangular Grid algorithms. The most significant finding is this developing method gave some early potential results that could form future solutions in the multi-objective optimization approach for the WSN designing.

KEYWORDS
Wireless Sensor Networks, Multi-Objective Optimization, Particle Swarm Optimization, Communication Path, Minimum Cost, Fault-Tolerant

1 Introduction
The fourth stage of the technological revolution or often called the four-point-zero industrial revolution is characterized by the usage of information and data as the main reference. Data availability is a guarantee of the success of an activity. Retrieval of large data in a short time can be interpreted as having to use an automated, connected, and real-time system. With these demands, the needs of the communication network become quite complex with various conditions. Meanwhile, not all regions have adequate communication infrastructure. To meet this gap, WSN technology is expected to become an alternative solution.

WSN devices often are nano-computers that are equipped with power supply modules, one or several sensors, and wireless communication modules. This device is commonly referred to as a node. This node can function as a sensor node, to retrieve and send data, as a relay node, to receive data from the sensors and forward the data to the sink directly or through other relays, as a storage node in the form of a relay with the ability to temporarily accommodate data, and as a sink/gateway node to store and process all data and at the same time as a gateway to the external networks.

But on the other hand, the limited ability of nodes ultimately raises various problems and challenges. Limited capabilities, resources, communication channels, and low signal strength are some of them [1-2]. With different node functions and single-
tiered or multi-tiered communication options, planning WSN to date are common problems that have not been fully resolved.

Problems in WSN planning are generally related to the efficient use of resources as stated in many previous studies. Efficiency, in this case, can be in the form of determining a location of nodes [1-6], the function of a node in the network [7-9], path selection and communication scheme [10-13], coverage areas [14-15], resource management [16-19], and the types of objects monitored [20-21]. Most of those researches provide optimization by considering only one aspect or perspective. In fact, besides relying on aspects that are considered the most influential, it would be better if the optimization processes were a gradual process and a combination of various interrelated aspects.

2 Problems

Different from some previous studies, this research tries to develop an optimization method in the process of designing and deployment of fixed WSN. Optimization is done by considering various things that might occur in real conditions, such as the communication range of devices, the presence of obstacles, communication schemes, and resource availability. The method, inspired by multi-objective optimization combines several aspects and objectives in an optimization process. This research aims to optimize coverage, connection, and communication between devices. The problem formulation of this research is then focused on the formation of an optimization method. As Particle Swarm Optimization (PSO) algorithm has a good exploration methods, and shows a better result compare to weighted sum, this developing method is based on it [22-23]. PSO algorithm is used to determine relay placements and connections. With the right placement, it is also expected to have an optimal connection. Optimal in this case, by considering the location, coverage area, number and function of the devices, network topology, and the efficiency of communication between devices.

3 Related Works

There are two main reasons for the growth of various studies related to optimization in WSN. First, because there are so many usages of WSN in various fields of work. And the second is the fact that WSN technology is still very limited in power supply, computing capability, and communication range. These two contradictions lead to a large optimization effort. Many of them are optimizing resources especially energy, device placement, topology, devices function, communication lines, channel capacity, and optimizing data flow. And it is also possible to consider for optimization of other conceptual and operational aspects.

Misra in [18] saw an unequal use of energy resources on each device and caused the network to be disconnected because the relay could no longer operate. To overcome this problem, Misra focused on establishing a method for the placement of relay devices. Assuming each relay is equipped with a power collector such as a solar panel, the relay must be in a place with lots of suns. The optimal solution is obtained using graph theory by minimizing the cost and number of relays used.

In [14], Yang and Chin consider a relay placement with the amount of solar energy collected and the number of objects monitored. To minimize the number of relays, Yang and Chin formulated a Mixed Integer Linear Programming (MILP). Then added a greedy heuristic algorithm to make sure that each relay connected to the sink directly or through others relay. This hybrid method is then called Greedy MILP or GMILP.

Considering the important role of relays in WSN communications, Ma [1] analyzed the influence of the location of a relay in correlation with the required energy. This research is divided into two parts. The first part analyses the range of relays, by applying Geometric Disc Covering (GDC). While the second part is focused on ensuring that each sensor can find the most appropriate relay by running a modified Greedy Search Algorithm called Local Search Approximation Algorithm (LSAA). The algorithm proposed as a solution in this study can be described as follows: (a) Merging sensor nodes into separate groups based on location and running the LSAA algorithm to get the minimum set of each group. (b) The most optimal LSAA output is taken to ensure that each group can be connected using an algorithm called the Relay Location Selection Algorithm (RLSA).

The idea of Ding and Yousefi’zadeh in [11] was dividing the area into hexagonal cells, with the edge sizes equal to the sensor communication radius. The layout and distribution of hexagonal cells are expressed in two-dimensional coordinates which are also hexagonal and referred to as HCS, Hexagonal Coordinate System. Sensors are assumed to be scattered between cells form a cluster where each cluster is connected to a relay. And relays in each of these clusters are connected to other relays that are closer to the sink. The main problem is the optimal placement of minimum additional relays to ensure that each relay at each cluster can be connected to the sink. The authors solve the problem by calculating the minimum spanning tree between relays using distance matrix and vector operation.

The following research [10] attempts to avoid communication failures by establishing alternative pathways between sensor and sink. Khoufi et al. focused on the solutions to allow all sensors to connect to the sink by placing as a few additional relays as possible to build two disjoint communication paths. This method divides the area into virtual triangular parts called triangular grids. Sensors and relays are assumed to be located at triangular points. A relay can be used together by several sensors. The formation and selection of the shortest path will determine the relay, while the rest will be eliminated. The network formed by this method is then compared with the results of two other methods, Straight Line and Steiner Point. The results are compared to several different conditions, such as area size, number of sensors, the presence of obstacles, and the formation of alternative pathways.

Communication inefficiency can be caused by the presence of obstacles or the quality of communication lines that are indeed inadequate. These trigger many data retransmissions from sensors and relays to the sink. It could also be due to the density of the sensors installed, raising interference, and further
reducing the quality of communication. Optimization to reduce the number of relays and data packets [12] was carried out by Nikolov and Haas using the RePlace algorithm. The idea of this algorithm is derived from the development of convexity in certain cases in the function of calculating the communication costs with the number of sensors and relays that have been determined. RePlace is an optimization idea derived from the greedy search heuristic method. This study tries to find a correlation between the addition of relays so that costs become smaller and network performance becomes better. Cost and data flow calculations are solved using graph theory.

4 Research Idea

As the problem formulation of this research is focused on the optimization method, below is the description of it. The PSO algorithm is used to determine relays placement and connection. Then followed by the shortest path calculation based on graph theory.

The first part of this optimization is related to relay placement problems. With sensors placed in the grid area randomly, the position of the relay is determined through several iterations to get the location that is considered optimal based on the number of relays, number of sensors connected, and coverage area. The output of this step is a collection of optimal relay location coordinates with a minimum number in a predetermined area.

The second step aims to design the network topology, by establishing connections from several potential relays. In this second step, the topology is assumed in the form of a tree with a sink node as the root. The objective of optimization at this stage is to choose the optimal communication path from all available paths, so that data can be sent from the sensor to the sink efficiently. The result of the second step is a network topology design with the condition that all sensors are connected to the relay and each relay has a communication link to the sink.

The third optimization is done to find an alternative path from a sensor to the sink, by forming a new path that is disjoint from the previous path. If necessary, additional new relays are carried out during this process. The main purpose of optimization in this step is to reduce the possibility of data packets retransmission because of a failure either on a relay or along the communication path.

5 Research Method

In this research, several assumptions need to be ascertained related to the area, connection between nodes, communication requirement, and mathematical approach. As an initial configuration, the area where WSN will be implemented is simplified as a square and divided into small triangular grids. The grid size is adjusted according to the relay communication radius. Triangular grids are used with consideration of its flexibility, offers the smallest overlapping area, and requires the lowest number of nodes [24]. This grid model ensures wide coverage and covers almost all areas.

Although there is no certainty that this model is the most optimal one, it more than enough as a pre-optimization placement method. This area modeling can be seen in Figure 1. Red dots define the possible placement location of sensors while relays are scattered randomly in the area and may occupy any dots.

With the communication radius expressed as \( r \) and \( R \) for sensor and relay consecutively, where \( R \geq r \), a sensor \( s \) is said to be connected and able to transmit the data to a relay \( r \) if the Euclidian distance between the two nodes \( d_{sr} \) is less than or equal to the communication radius \( r \), and formulated mathematically as:

\[
\delta_{sr} = \begin{cases} 
1, & d_{sr} \leq r \\
0, & d_{sr} \geq r
\end{cases}
\]  

(1)

Figure 2 shows some of the possibilities that can occur between sensors and relays. Sensor \( s_1 \) is in the area but there is no relay. Mathematically, the sensor \( s_2 \) has Euclidian distance as \( d_{s_2r_1} \geq R \). So, it is unable to detect and undetected by all existing relays. Sensor \( s_3 \) located in the range of the communication radius of the relay \( r_1 \), \( d_{s_3r_1} \leq R \) but cannot be connected because the Euclidian distance from the sensor \( s_3 \) to the relay \( r_1 \) exceeds the communication radius, \( d_{s_3r_1} \geq r \). And the sensor \( s_2 \) is within the relay and sensor communication radius, which is \( d_{s_2r_1} \leq r \). This condition allows the sensor \( s_2 \) to be connected to the relay \( r_3 \).

The next step is to ensure that relay \( r_1 \) can forward the data to the sink. Considering the communication radius of relay \( R \), relay \( r_1 \) must be connected to other relays, either \( r_2 \), \( r_3 \) or \( r_4 \). If relay \( r_1 \) is not connected to any relay, the position of other relays must be shifted until at least one of them is connected to relay \( r_1 \). From Figure 2, relay \( r_1 \) is connected to relay \( r_2 \) and then via \( r_3 \) to connect to \( r_4 \) and finally to the sink. Assuming the communication radius of the sink is equal or greater than the communication radius of relay \( R \). Now, related to the relay’s communication capabilities, the Euclidian distance between two relays must be less than or equal to \( R \):

\[
\delta_{rr} = \begin{cases} 
1, & d_{rr} \leq R \\
0, & d_{rr} \geq R
\end{cases}
\]  

(2)

Figure 1: Triangular Grid Area Modeling
As stated earlier, the main objective of this study is to build an optimal WSN by ensuring that each sensor that is spread in the area able to send its data to the sink through two different paths with the least number of relays. To get this, the first thing to do is to ensure that each relay can connect as many sensors as possible by placing the relay in an optimal position by considering the position of sensors around it. However, it should be noted that computing and communication capabilities on the other hand limit the number of sensors that can be connected to a relay. Assuming $C_{\text{max}}$ is the maximum connection capacity of a relay and $s_i$ is the $i$th and the sensor of total $n$ sensors installed in the area. Then the minimum number of relays $N_{\text{min}}$ can be formulated as:

$$N_{\text{min}} = \frac{1}{C_{\text{max}}} \sum_{i=1}^{n} s_i \quad (3)$$

This $N_{\text{min}}$ is the initial value of optimization and is used as one of the parameters of the PSO algorithm. Other parameters are determined based on the results of the research conducted by Hu [25]. The PSO algorithm is then run in several different areas using scenarios with different numbers of sensors. The simplified algorithm can be seen in Figure 3. The next to optimization is to calculate the smallest path cost value from each sensor to the sink. First, the collection of sensor locations is defined as the set $S$, and the location of all relays of PSO algorithm output $P_{\text{opt}}$ is defined as the set of relay positions $P$.

All possible connections of the sensor and relay are expressed as the set $E$.

$S = \{s_1, s_2, \ldots, s_n\}, \text{s.t. } 1 \leq i \leq n \quad (4)$

$P = \{p_1, p_2, \ldots, p_m\}, \text{p.t. } 1 \leq i \leq N_{\text{opt}} \quad (5)$

$E = \{e_1, e_2, \ldots, e_j\}, \forall \delta_{\text{gr}} = \delta_{rr} = 1 \quad (6)$

Figure 2: Node Coverage and Connectivity

With these three sets, $S$, $P$, and $E$, a common graph $G$ with an optimal solution can be expressed as:

$$G = (S, P, E) \quad (7)$$

Figure 4 shows an example of a simplified graph-based topology constructed from simulated PSO output. This model consists of five sensors spread over a rectangular area. The sensor and relay communication radiiuses are set to a value of $R = 2r$. Relay positions are adjusted according to the triangular grid coordinates, while the sink is placed at the origin. From this model, to ensure each sensor is connected to the sink, no less than seven relays are needed to form the set $E$. A WSN with a fault-tolerant feature requires $E$ that consist of around 15 relays.

From the set $E$ that contains the connection path between sensor, relays, and sinks, all possible connections are established. The choice of connection path is determined by calculating the cost of each path. The path cost component in this research is the composition of the weight of the relay and the connecting path. The weighting for each relay considers the distance to the sink and the number of devices connected directly to the relay.

Using Figure 4 as an example, there are three categories of relays at distance from the sink. They are relays that are connected directly, either through one or two intermediate relays. The more intermediate relays needed, the higher the weight of the relay. Relays $r_1$, $r_2$, $r_{11}$ and $r_{16}$ are connected directly to the sink and given a weight of 5. In the middle distance, there are five relays namely $r_3$, $r_5$, $r_8$, $r_{12}$ and $r_{15}$ with a weight value equal to 10. The last group of the farthest relay consists of $r_4$, $r_7$, $r_9$, $r_{10}$, $r_{13}$ and $r_{14}$ with a weight of 15.
The second element that determines the weight value of a relay is the number of devices connected directly. Each device connection will add a weight value of 1 point. For example, relay $r_1$ gets an additional weight of 3 so that the total weight value $w(r_1)$ for relay $r_1$ is equal to 8, and so on for all other relays.

The edge weight value is calculated using the two-node weight values connected by the formula:

$$w(n_1n_2) = \frac{w(r_1) + w(r_2)}{2}$$ (8)

For example, the edge weight between the two relays, $r_5$ and $r_{11}$, that was $w(r_5r_{11})$ is equal to 14.5. So, a sub-graph $H$ of $G$ is said to be optimal if each sensor can transfer data to the sink either directly or through several relays. The communication cost then calculates by formulation:

$$w(H) = \sum_{(s,r) \in E[H]} w(s,r)$$ (9)

Models of the results of the second and third phase optimization are shown in Figure 5. Approximately 25 paths remain after the calculation. The closest sensors $s_1$ and $s_4$ are connected to the sink via two relays, while the other sensors go through three intermediate relays. The minimum additional weight value for most relays is 2 and the maximum additional 5 for the relay $r_8$.

4 Results and Discussion

Figures 6 and 7 are the graphical result models of the developing method. Simulations were assumed conducted in three areas with three different sensor locations. In each area, ten sensors are placed in a triangular grid pattern. With ten sensors spread out, each area has two different topologies, single-path, and multi-path. The first optimization runs on single-path network topology. That is to ensure all sensors are connected to the sink through a communication path. The second optimization runs over a multi-path topology with the existence of another disjointed path. In this topology, all sensors are connected to the sink through two different paths.

The optimization results are then compared using two different algorithms, Steiner Point and Triangular Grid, using the number of relays, maximum path, and communication cost. The number of relays and the maximum path is used to show the coverage area. The maximum path in combination with communication cost is used to define cost efficiency. Two disjointed paths are to demonstrate the fault-tolerance ability. The Steiner Point is used as a comparison to evaluate the maximum path while the Triangular Grid is used to evaluate the relay placement.
Triangular Grid. For example, in simulation three the PSO has 47 relays and Triangular Grid has 44 installed relays. For the maximum path value, from simulation one and three, it can be stated that both the PSO and Triangular Grid have similar values. In simulation one, the PSO has 10 value as its maximum path, while Triangular Grid has 9. But in simulation three, the PSO has 9 and Triangular Grid has 10 for the maximum path. From the communication cost, the PSO algorithm has a better value compared to Triangular Grid. It is 17.0 compared to 20.5 in simulation one and 37 compared to 37.5 in simulation two. This results is the outcome of PSO characteristic that has a broad searching space.

5 Conclusion

From all the early simulation results, it can be stated that the PSO algorithm has created a potential solution. In combination with graph theory, it can be used to form a multi-objective optimization approach. The use of the proposed method probably can reduce the number of relays installed. Also, when examining the maximum path results, the proposed method could have a value close to the Steiner point. The value of communication costs from this developing method tends to be like or even lower than Triangular Grid. Lastly, the proposed PSO-based method also can bring an increased fault-tolerance ability for the network, with a disjunction path. With these initial results, PSO-based heuristic optimization can be used further as a basis for optimizing methods in WSN design. Also, it has more flexible characteristics on node placement and could be more realistic for actual implementation. Therefore, it would be worthwhile to do further exploration, evaluation, and analysis of this method, including investigating the effect of the improved calculations.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Noor Hidayat, Dr. Wuryansari Muharini Kusumawinahyu and Dr. Isnani Darti, Lecturers of Mathematics Department of Universitas Brawijaya, whose constructive and insightful views and comments helped improve the presentation of this paper.

REFERENCES