α-Fraction First Strategy for Hierarchical Model in Wireless Sensor Networks

Jeng-Shyang Pan[1,2,3], Lingping Kong[2], Tien-Wen Sung[1], Pei-Wei Tsai[4[1]], Václav Snášel[5]

[1] Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, China
[2] Innovative Information Industry Research Center, Harbin Institute of Technology Shenzhen Graduate School, China
[3] Department of Information Management, Chaoyang University of Technology, Taiwan
[4] Department of Computer Science and Software Engineering, Swinburne University of Technology, Australia
[5] Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, Czech Republic
jengshyangpan@fjut.edu.cn, konglingping2007@163.com, twsung@foxmail.com, peri.tsai@gmail.com, vaclav.snasel@vsb.cz

Abstract

Energy hole refers to the critical issue near the sinks for data collecting, this problem effects the lifetime of wireless sensor network to a great extent. Frequently data forwarding from distributed sensors to the sink will speed up the energy consumption of the sensors near the sink. This circumstance shortens the lifetime of the sensor network. In this paper, an α-fraction first strategy was proposed to build a hierarchical model of wireless sensor networks that concerning the energy consumption. The model mixes the so-called relay nodes into the network for transmitting and collecting data from the other sensor nodes. We studied the Farthest First traversal and Harel methods, then combined the proposed α-fraction first strategy with the two methods respectively. Three algorithms of FF+Fr(α), HD+Fr(α), and HL+Fr(α) were designed for determining the relay nodes in sensor networks. The algorithms can be used to construct a two-tier sensor network with fewer relay nodes than the results of the Farthest First traversal and Harel methods. The proposed strategy also could be used with any other algorithms that regarding for choosing one of many options. The simulation results show that our proposed algorithms perform well. Thus, the network lifetime can be prolonged.

Keywords: Wireless sensor network, Farthest-first strategy, Hierarchical network, Nearest-first strategy

1 Introduction

Wireless sensor networks (WSNs) actually is another way of interpretation of wireless sensor and actuator networks (WSAN), this system [1] employ many spatially deployed sensors to monitor environmental factors within a physical area such as temperature, sound, pressure condition, etc. They are also used to cooperatively and wirelessly transfer the collected data to a sink node via network routing. The sink node is the controlling and processing center of the network, for example, the coordinator of a ZigBee sensor network [2]. One of the advantages of wireless sensor networks is their ability to operate in extreme or unsafe environments, in which manned operating schemes are risky, inefficient and sometimes infeasible [3]. Sensors are thus distributed randomly in the area by an uncontrolled means, such as helicopter. Hundreds or even thousands of sensor nodes are then involved in transferring and communicating with the sink node. Designing and operating such large networks requires scalable architectural and management topology [4]. At the same time, energy-aware algorithms become a vital factor in extending the lifetime of such networks [5-6]. Grouping sensor nodes into a two tier architecture strategy is a commonly used strategy. An example would be an internet network in which optical cables are used to transmit information to servers and important points or computers, called an optical network. In facility location problem, one of the popular ones is k-center problem. Many studies have suggested a 2-approximation algorithm (A solution which outputs value is no more than 2 times the optimal value is called a 2-approximation algorithm where the solution is for a minimization problem and runs in polynomial time). This study thus focuses on developing a method to schedule these base stations and ensure the minimum possible k value.

Rosenkrantz, Stearns and Lewis first used the farthest-first method [7] in connection with heuristics for the travelling salesman problem. Later, the same sequence of chosen points was popularized by Gonzalez [8]. The farthest-first traversal can also be

[1] This work was done while the author worked at Fujian Provincial Key Laboratory of Big Data Mining and Applications (Fujian University of Technology).

*Corresponding Author: Lingping Kong; E-mail: konglingping2007@163.com

DOI: 10.3966/160792642018111906009

used in many fields associated with location problems. In addition to the farthest-first traversal approach, others have also used the nearest-first traversal method; it chooses nodes nearest other nodes from a group of nodes, rather than the farthest ones. In fact, the two methods share the same disadvantage in finding the *k*-center solution. When choosing either the farthest or the nearest node, only half a side of uncovered nodes is examined. For example, in Figure 1, using Farthest First method to choose *k* base station from 8 nodes. Assume the first node randomly chosen is 1 (base station set is {1}). Then, the next chosen node must be 8 (base station set is {1, 8}), which has the farthest Euclidean distance from 1. Removing those nodes that they are already dominated by {1,8} base station, then only nodes 4, 5 and 7 are left. The third node added into set will be 4 as the Euclidean distance reason. From here, it can be seen that node 2, which is the neighbor of node 4, is already covered by the former adding base station of {1}, so the new adding station 4 only works for its left side covering. It commonly occurs that most base stations only cover half of an area, while the other half of the area is usually either the boundary or covered by previous base stations. A method for choosing these base stations is the primary aim of the proposed α-fraction first method.

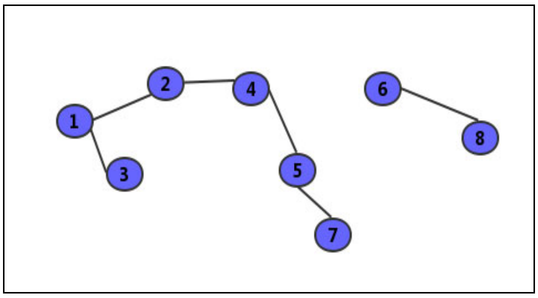

Figure 1. An 8-node simple structure for farthest-first process

We arranged the structure of the paper as showed: Section 2 lists related studies; Section 3 introduces the proposed α-fraction first method; Section 4 presents the simulation results; and Section 5 presents the conclusion.

2 Related Works

2.1 K-center Problem

Here it explains a simple definition of *k*-center problem [9]. In a unweighted graph *G* of given which includes *n* vertexes (nodes) and *m* subsets with up to *k* elements (service centers), one of the subsets will be selected to be a service center set and any of the nodes in *G* must be assigned to its nearest service center in the selected set as a client. The objective function is to narrow down the maximum distance from a node to service center that it belongs to. It is as follows.

$$f_{obj} = \min_{1\le i\le m}\{\max_{1\le j\le n}\{\delta_G(v_j, S_i)\}\} \quad \textbf{(1)}$$

where $\delta_G(v_j, S_i)$ denotes the distance in graph *G* from a node j $(j \in v)$ to a service center i $(i \in S)$.

Many algorithms for solving variants problem about *k*-center have been researched recently [10-11] For example, Du et al. [12] examined a kind of *k*-center problem with more center nodes constrained and the increment is that the boundary of those centers should be a convex polygon. And the authors create a round 2.6-approximation algorithm. In the meanwhile, they also proved that any increment algorithm can't guarantee a solution with competitive ratio of better than 2 for this problem in polynomial time. A different kind of *k*-center problem with connected constrain of internal nodes called CkC is considered by Liang et al. [13], actually this incremental version was first presented by Ge et al. [14], they give a *6*-approximation scheme for the CkC problem. Andreas [15] has researched many variants of *k*-center problem and its generalizations. The author also showed how to obtain a strategy combining the help of parameters *k* and *h* with a ratio below factor of 2 algorithm, where *k* is naturally the amount of vertexes centers and *h* is the dimension element. The point is the Dominating Set problem is the specific version of *k*-center problem, the authors noted that in this paper. Chechik [9] examined the capacitated *k*-center problem. He studied the new version of this problem with fault-tolerant mechanism, in which one or more centers could fail at any moment.

Authors [16-17] studied the cases using nearest-first swarm (NFS) to assign the available task to its closest robot, then analyze the influence of interference between multi-robot systems. For improving the performance of clustering technique by actively selecting meaningful pairwise constraints, Basu et al. [18] use the farthest-first traversal scheme to present a clustering framework and pairwise constrained method, which could work well even in the big datasets and easily scalable to high dimensional data. Besides, the author also points that the *k*-center problem solution [19-20], which is efficient approximation, could be yielded by farthest-first traversal. The farthest-first concept is also used to analyze data in data mining. There are also researches about the optimization clustering with fixed cluster radius and resulting in the uniform clusters, and the study [20] proposed a fast and greedy algorithm by applying the farthest-first strategy for this kind of clusters. In [21], Garima compared various clustering algorithms, including the farthest-fist, K-means, expectation maximization and non-partitioning-based, density-based, etc., algorithms. The results show that the farthest-first took the least time to form clusters for all three datasets, while expectation maximization took the longest time. Sharma et al. [22] used the data mining tool Weka to study various clustering algorithms. Finally the authors conclude that farthest-point method was one of the appropriate algorithms for users. Deepshree and Dharmarajan [23-24] used the farthest first heuristic method to perform reorganization and lung cancer data analysis.

2.2 Energy Model

In this paper, it was assumed that the sink node is in the center of an field, which can connect and communicate to all the relay nodes directly. The relay nodes are responsible for collecting the data from the distributed sensors and forwarding them to the sink. Those sensors sense the interesting events and report the events to the relay nodes. Here lists some structure introductions regarding our network model:

(1) A periodic event collecting happens where event alarm occurs and one unit data will be transmitted in one slot.

(2) Sensor nodes are uniformly and randomly deployed. The locations of the relay nodes are chosen from the original place of the deployed sensors. The total number of the sensor nodes and relay nodes are constant.

(3) When a periodic event happens, the closest sensor near the event source will respond it and restrain from the act of other nodes. If two sensor nodes simultaneously sense the event, the realy node will receive the data twice.

(4) The relay nodes will forward the collecting data to the sink every four slots.

(5) The sink node has unlimited energy resource. Thus we ignore the energy consumption of the sink.

There is a similar network model used in paper [25], and in our model, we only compute these energy consumption: sensing, receiving and transmission. This paper follows equation 2 to compute the energy, m is one unit data.

$$\begin{aligned} &E_{sense} = \alpha_1 m \qquad \alpha_1 = 60\times 10^{-9} J/bit \\ &E_{rx} = (\beta_1 + \beta_2 r^2)m \quad \beta_1 = 45\times 10^{-9} J/bit \\ &\beta_2 = 10\times 10^{-12} (J/bit/m^2) \\ &E_{Rx} = \gamma_1 m \quad \gamma_1 = 135\times 10^{-9} J/bit \end{aligned} \tag{2}$$

3 α-Fraction First Strategy

3.1 α-Fraction First

The combination of α-fraction first strategy with the farthest-first traversal (FF) and Harel [26] methods is presented in this section noted as FF+Fr(α) and Harel+Fr(α), In this method, two types of distance are used: Euclidean distance and weighted distance computed by Dijkstra equation. There also two kinds of Harel method: Harel Line (HL), which uses Euclidean distance, and Harel Dijkstra (HD), which uses Dijkstra distance. Fortunately, the distance formula mode has no effect on the combination process. They use the same process either for Harel Line or Harel Dijkstra. The α-fraction first strategy attempts to select a node with α-fraction distances away from the last adding node (or nodes) as a station instead of the farthest or nearest away node; this increases the probability of nodes with full side uncovered area being selected as service stations. The Harel method is actually a variant of farthest-first traversal. It was used to select M-dimensional pivots for high dimensional embedding graphs. The Harel method also starts with one randomly chosen node. Then each successive station is as far as possible from the last chosen one. As mentioned in the introduction of this paper, there are flaws involved with the farthest first traversal method. Each successive selected station only covers half a side of uncovered nodes, which results in more boundary nodes being chosen as stations.

Here is the symbol description. A node set S with n nodes, and service station set C. Set C is empty initially, and the last node becomes a station marked as p, a fraction parameter denoted α. $List\{x \in S \mid \rho(x, C)\}$ is a collection of nodes in order based on their Euclidean distance values from the set C. The function $\rho(x, C)$ is the distance from the node x to its closest service center in set C; Symbol $List\{\ \}_i$ acts as the i-th element in $List\{\ \}$. $List\{\ \}$ could be stored as a linked list when choosing z. The size of set C is the number of service stations, and it is the final result. The process for the combined α-fraction first and farthest-first method show as Table 1.

Table 1. Pseudo-code of FF+Fr(α)

Pseudo-code of FF+Fr(α)
First randomly pick any node $z \in S$ and set $C = \{z\}$.
Remove node z from S. Set node z as p.
Remove any node in S covered by p.
While $S \neq$ null {
$z = List\{x \in S \mid \rho(x, C)\}_{\alpha \times size(List\{x \in S \mid \rho(x, C)\})}$
$C = C \bigcup \{z\}$
Set p as node z. remove node z from S.
Remove any node in S covered by p }
End While
Output set C.

Next, the α-fraction first is combined with the Harel method as in Table 2. An important point is that the original Harel process selects the pivots for creating high-dimensional data, and does not skip nodes already covered as candidate station operation. This operates against the rules of finding the minimum number of stations. Therefore, in this paper, the Harel method removes the probability of covered nodes being chosen as potential stations. In Table 2, the station set is T, and it is the final result.

Table 2. Pseudo-code of Harel+Fr(α)

Pseudo-code of Harel+Fr(α)
First randomly pick node $z \in S$ and set $T = \{z\}$.
Remove node z from S. Let $p = z$.
Remove any node in S covered by p.
While $S \neq$ null.
$z = List\{x \in S \mid \rho(x, p)\}_{\alpha \times size(List\{x \in S \mid \rho(x, p)\})}$
$T = T \bigcup \{z\}$
Set node p as z. remove node z from S.
Remove any node in S covered by p.
End While
Output set T.

3.2 Fitness Function

In order to substantially increase the chance of finding a better solution, a well-constructed fitness function is presented [27]. Symbol $\rho(x_i, C)$ is the distance between the node x_i to its closest service center in the set C generated by FF-Fr(α). Similarly, $\rho(x_i, T)$ is the distance between x_i to its closest service center in the set T generated by Harel+Fr(α). S is the set of n nodes. The function $d(x_i, S_{sink})$ represents the distance between the node x_i and the sink node of the WSN. Coefficient $S\alpha_1, \alpha_2, \alpha_3$ and α_4 are constant values.

$$f_{fit} = \begin{cases} \alpha_1 \cdot |C| + \alpha_2 \cdot (\sum_{1 \le i \le |S-C|} \rho(x_i, C) / \sum_{1 \le i \le n} d(x_i, S_{sink})), \\ \qquad \text{for FF} + \text{Fr}(\alpha) \\ \alpha_3 \cdot |T| + \alpha_4 \cdot (\sum_{1 \le i \le |S-T|} \rho(x_i, T) / \sum_{1 \le i \le n} d(x_i, S_{sink})), \\ \qquad \text{for Harel} + \text{Fr}(\alpha) \end{cases} \quad (3)$$

3.3 Number of Distance Calculation

During the process of locating a station, the distance between nodes will be calculated many times. What is the specific frequency of each method? This section lists the number of distance calculation (NDC) for three methods. Actually, combining the farthest-first or Harel methods with the fraction method does not increase distance computing times.

The following Table 3, Table 4, Table 5 are lists of parameters and results for FF-Fr(α), HL+Fr(α) and HD+Fr(α) in number of distance-computing calculation respectibely. In fact, the program will do some pruning of the computation. For both methods and their combination forms, the first step of randomly picking one node would run n times, and n is the total number of nodes. This is one time in the distance-computing times in the equation brace. The k is the output of each method, which is the number of stations. In Table 5, the parameter m denotes the number of edges among sensor nodes.

Table 3. NDC for Farthest First

Nodes - n	100	200	300
Stations - k	16	28	44
t_{dc}	1499100	14697200	87236700
Nodes - n	400	500	600
Stations - k	63	110	209
t_{dc}	314551600	1279932500	1717210304

Table 4. NDC for Harel Line

Nodes - n	100	200	300
Stations - k	15	25	38
t_{dc}	129500	900000	3119100
Nodes - n	400	500	600
Stations - k	55	91	186
t_{dc}	8046000	20452500	56277000

Table 5. NDC for Harel Dijkstr

Nodes - n	100	200	300
Stations - k	17	29	46
Edges - m	523	1197	1919
t_{dc}	6531400	59176000	243945000
Nodes - n	400	500	600
Stations - k	64	106	189
Edges - m	2245	2064	1284

Farthest First:

$$t_{dc} = n \times \{(n-1) \times 1 + (n-2) \times 2 + \cdots + (n-k+1) \times (k-1)\}$$
$$= n\{n \times \frac{k \times (k-1)}{2} \frac{(k-1) \times k \times (2k-1)}{6}\} \quad (4)$$

Harel Line

$$t_{dc} = n \times \{(n-1) + (n-2) + \cdots + (n-k+1)\}$$
$$= n \times \{n \times (k-1) \frac{k \times (k-1)}{2}\} \quad (5)$$

Harel Dijkstr

$$t_{dc} = n \times \sum_{1 \le i \le (k-1)} (m + n - i) \cdot \log_2 (n - i) \quad (6)$$

4 Experiments

4.1 Relay Node Size

A two dimensional space 200×200 units in size is used as the simulation environment. In the center position of this area is a sink node. A number of sensor nodes are randomly deployed in this environment, in groups of 100, 200, 300, 400, 500 and 600. Each node is equipped with the same sensing and transmitting devices. The goal is to locate k nodes out of n total sensor nodes as stations, and those stations could communicate with the sink node without a third party; this must be done with the minimum necessary k and optimal fitness value. In the simulation, the greater the fitness value, the better, and the smaller that k becomes, the better. The fraction factor α is set at 1/2, 1/3, 1/4, …, 1/9, 0 and 1. When the fraction factor α equals

1, it is the farthest-first method; it is the middle-first method if α is 1/2.

Table 6 to Table 8 show the results with ten different fraction parameters in a 100 to 600-node network. Here, if α is equal to 1, it becomes the original farthest-first traversal method. Unfortunately, its outcome is the worst of all the examined methods. For a 100-node network, the farthest-first traversal method needs 19 sites to construct stations. However, the lowest number of stations required is 13, achieved by FF+Fr(1/7) or FF+Fr(1/8). It costs 6 stations less than the farthest-first traversal method. The α-Fraction idea shows good performance in other network sizes. In a 600-node network, a decreased construction fee of 20 stations would economize a lot. The last Fraction factor α=0 actually behaves like the nearest-first method. This process also reveals its disadvantage. Both the farthest-first and nearest-first methods suffer from the fact that only some sensor nodes around one station are available. Nearly half of the neighbors of a node are already covered by the previous station. The combination results for the Harel and Fraction methods are listed in Table 7 to Table 8. The results, with the α value less than 1 and greater than 0, are always better than the opposite cases in terms of both station number.

Table 6. The number of stations generated by of FF+Fr(α)

Methods \ Nodes	100	200	300	400	500	600
FF+Fr(1)	19	29	47	67	110	209
FF+Fr(0)	16	28	44	63	107	201
FF+Fr(1/2)	14	24	37	56	97	191
FF+Fr(1/3)	14	24	36	57	94	189
FF+Fr(1/4)	14	24	38	57	97	195
FF+Fr(1/5)	13	24	37	56	95	194
FF+Fr(1/6)	14	24	38	56	94	193
FF+Fr(1/7)	13	24	37	56	98	195
FF+Fr(1/8)	13	23	37	58	97	192
FF+Fr(1/9)	15	23	38	55	97	195

Table 7. The number of stations generated by of HL+Fr(α)

Methods \ Nodes	100	200	300	400	500	600
HL+Fr(1)	15	25	38	55	91	186
HL+Fr(0)	15	23	36	52	92	173
HL+Fr(1/2)	14	24	35	52	89	177
HL+Fr(1/3)	14	23	35	50	87	175
HL+Fr(1/4)	13	24	36	52	89	179
HL+Fr(1/5)	14	23	36	52	89	179
HL+Fr(1/6)	13	23	35	52	90	179
HL+Fr(1/7)	14	24	34	52	90	181
HL+Fr(1/8)	13	24	35	52	90	180
HL+Fr(1/9)	13	23	34	52	91	179

Table 8. The number of stations generated by of HD+Fr(α)

Methods \ Nodes	100	200	300	400	500	600
HD+Fr(1)	17	29	46	64	106	189
HD+Fr(0)	16	27	41	60	103	191
HD+Fr(1/2)	13	23	33	52	90	178
HD+Fr(1/3)	13	23	35	52	91	181
HD+Fr(1/4)	14	23	36	53	90	192
HD+Fr(1/5)	13	23	36	52	90	192
HD+Fr(1/6)	13	23	35	53	89	188
HD+Fr(1/7)	13	24	33	52	91	189
HD+Fr(1/8)	13	22	35	53	91	185
HD+Fr(1/9)	14	23	34	53	89	187

4.2 Energy Consumption

Table 9 lists the data sensing effective ratio (*dnER*) value for the two different methods *Fb* and *Fr*. The value of notation *Fb* indicates the best result among the Farthest First strategy with *FF, HL,* and *HD*. The value of *Fr* indicates the best result generated by combined *Fr(α)* method. The data sensing effective ratio was defined as the value of the number of events occurred (*Rn*) in the surveillance environment to the value of the real number of packages for the events (*Pn*) sent by sensor nodes to relay nodes. The values of *Rn* is thirty times of the number of sensor nodes for 100-node to 600-node sensor networks. Each event will be discovered by sensors nearby within a circle radius of *R*. If there is no sensor in this area, the sensors within radius *2R* will be requested to work, and so on. It will stop the distance extension by multiples of *R* after the radius goes out of the largest sensing radius. Once there are sensors respond to the events, the other farther sensors will be suppressed. Thus, the responding sensors will transmit the collected data to the relay node. The lower ratio of the data, the more duplication packages were sent by the sensor nodes, and the more energy was consumed. The fist column indicates the number of sensor nodes (*TN*). As shown in the table, The *Fr* combination method got a higher sensing effective ratio than the method *Fb*.

Table 9. Packages comparison

TN	Method	dnER	TN	Method	dnER
100	Fb	85.76%	400	Fb	72.09%
	Fr	86.93%		Fr	73.60%
200	Fb	64%	500	Fb	71.05%
	Fr	66.25%		Fr	73.18%
300	Fb	68.68%	600	Fb	70.50%
	Fr	70.74%		Fr	72.32%

Table 10 shows the values of total energy consumption (*TotalE*) for each of the 100-node to 600-node sensor networks that transmitted 3000 to 18000 packages, that is to say, each node forwarded 30 packages on average. Figure 2 shows the comparison of the energy consumptions for sensing, receiving and transmitting activities. The notation *sE* sums the energy costs for sensing processes by sensor nodes, and *srE* records the receiving costs by relay nodes. The other notations *ssE* and *rsE* are the costs for sending messages by sensor nodes and relay nodes respectively. *TotalE* shown in the table indicates the total energy cost for all the operations. It can be found that *Fr* spends a less energy consumption than the *Fb* method.

The Figure 3 to Figure 8 show the two-tier sensor network with stations. The only one sink node is in blue. All the stations are shown in red. The small black nodes communicate with their upper level station. Each black node only has one upper station. Due to the large number of nodes, the sink node is set as invisible in the 400-node to 600-node networks. Figures 3 to 8 only list the original farthest-first traversal (FF), Harel Line (HL), Harel Dijkstra (HD) and the best of the Fraction first combination results. For example, in Figure 3, the first row depicts the network constructed by FF, HD and HL sequentially, and the second row shows the best results in combination Fraction first with FF, HD and HL, sequentially.

Table 10. The energy cost comprison

TN	Method	TotalE	TN	Method	TotalE
100	Fb	1873887	400	Fb	8572010
	Fr	1828898		Fr	8375043
200	Fb	4926667	500	Fb	1.067 E7
	Fr	4800849		Fr	1.0424 E7
300	Fb	6787450	600	Fb	1.2946 E7
	Fr	6531141		Fr	1.2574 E7

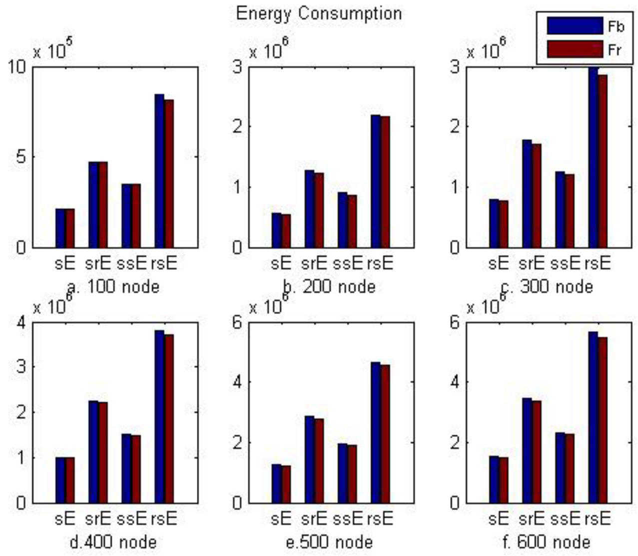

Figure 2. Energy consumption

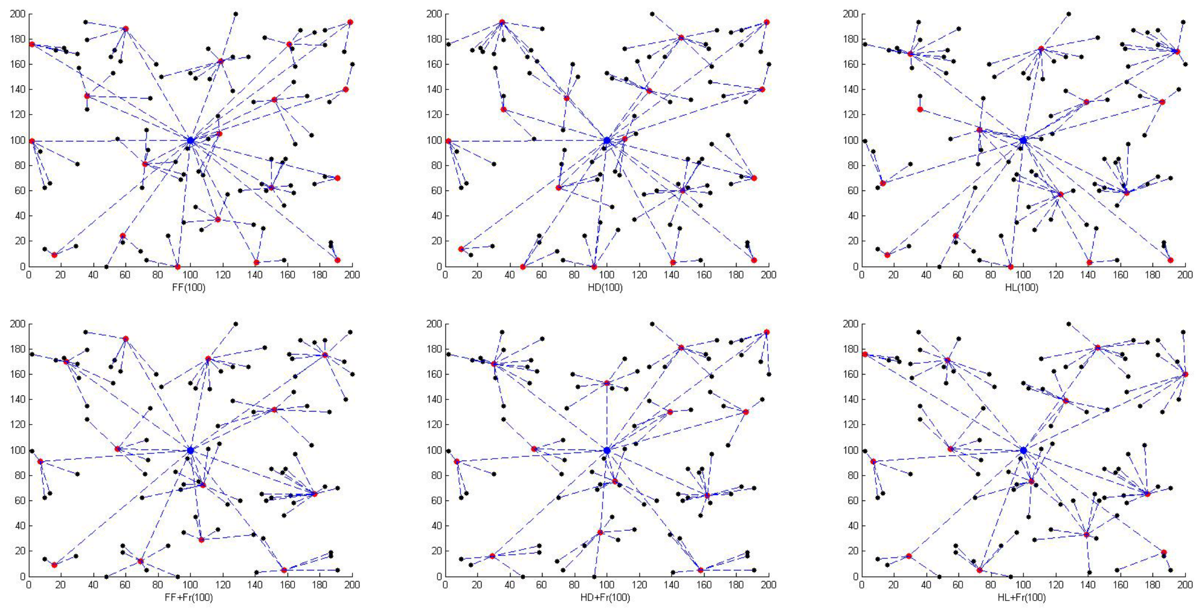

Figure 3. Comparison of 100-node network construction

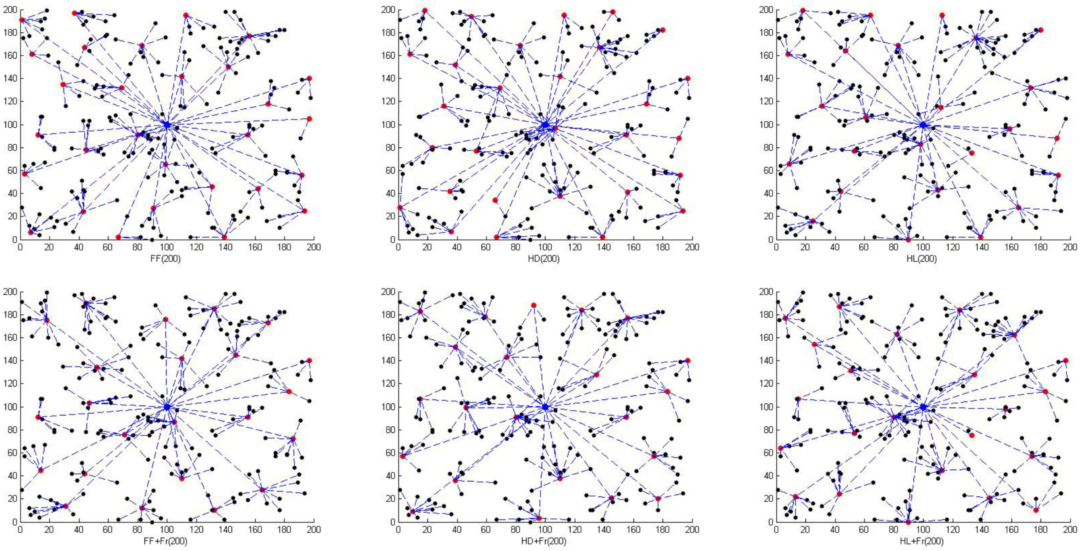

Figure 4. Comparison of 200-node network construction

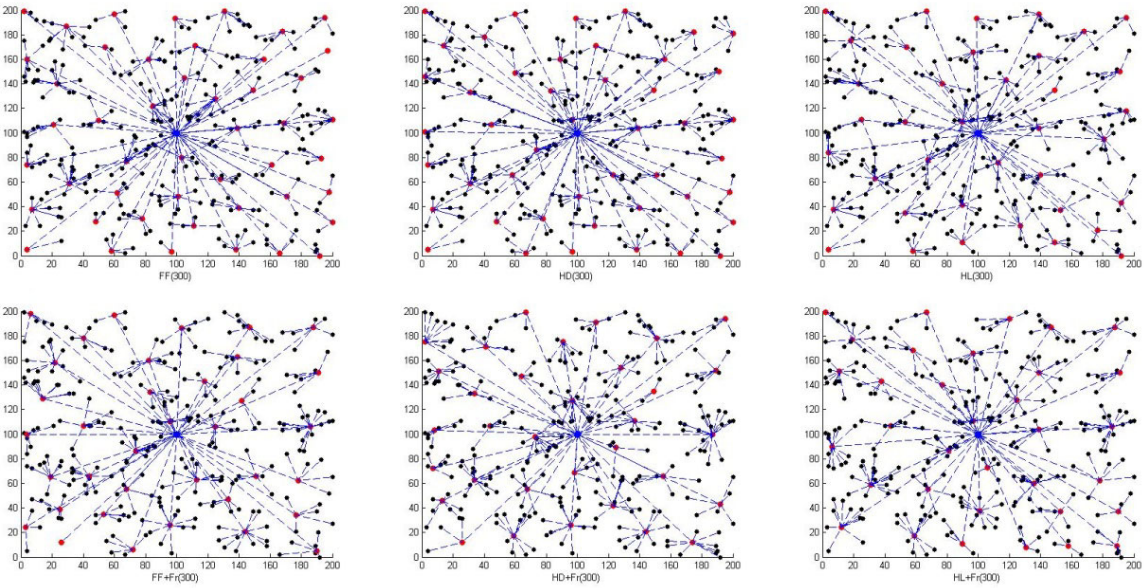

Figure 5. Comparison of 300-node network construction

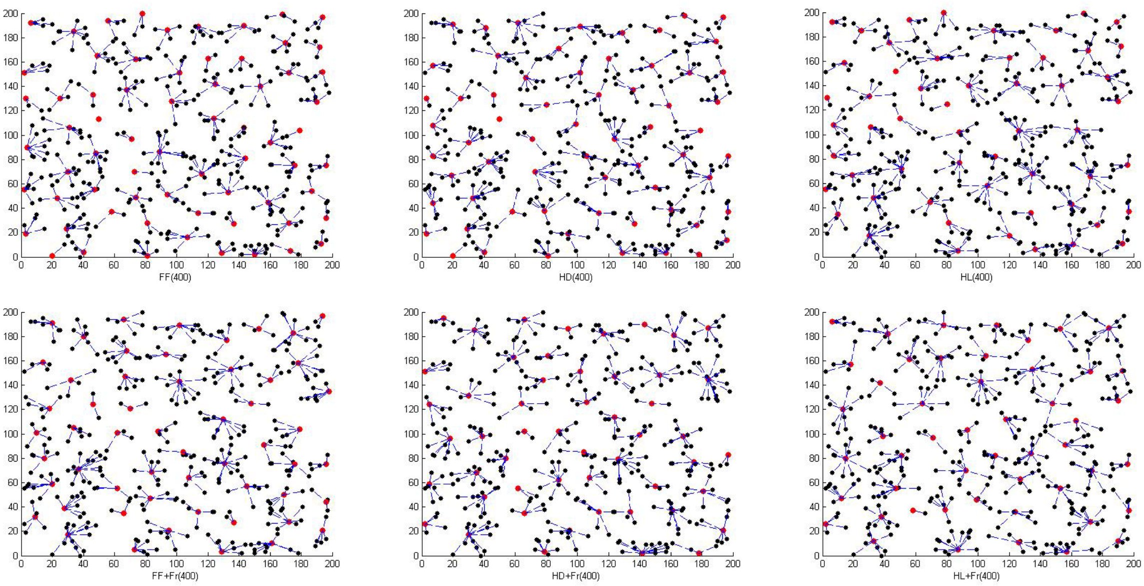

Figure 6. Comparison of 400-node network construction

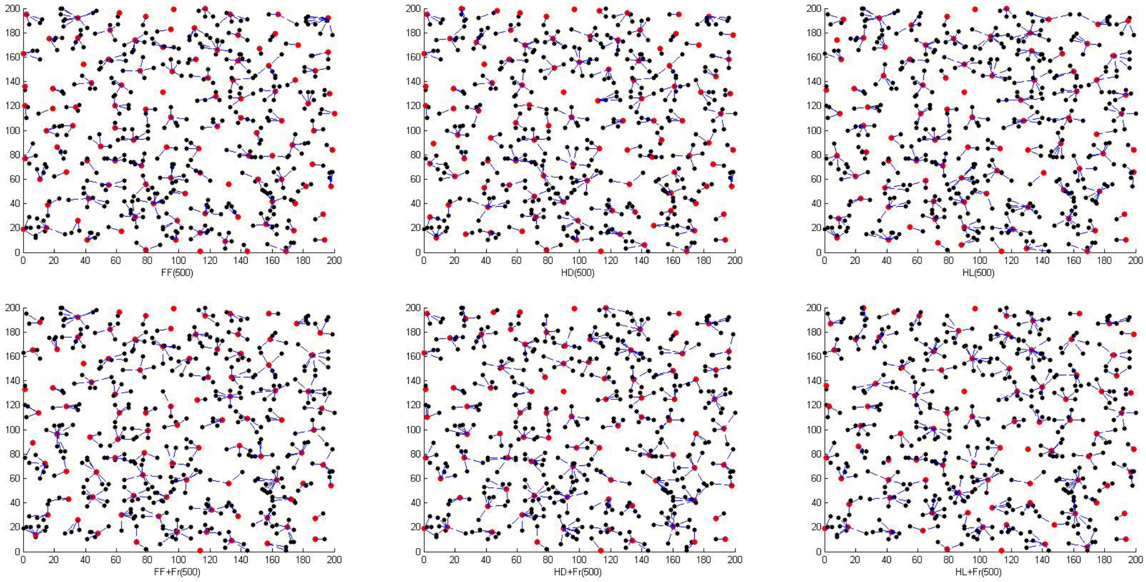

Figure 7. Comparison of 500-node network construction

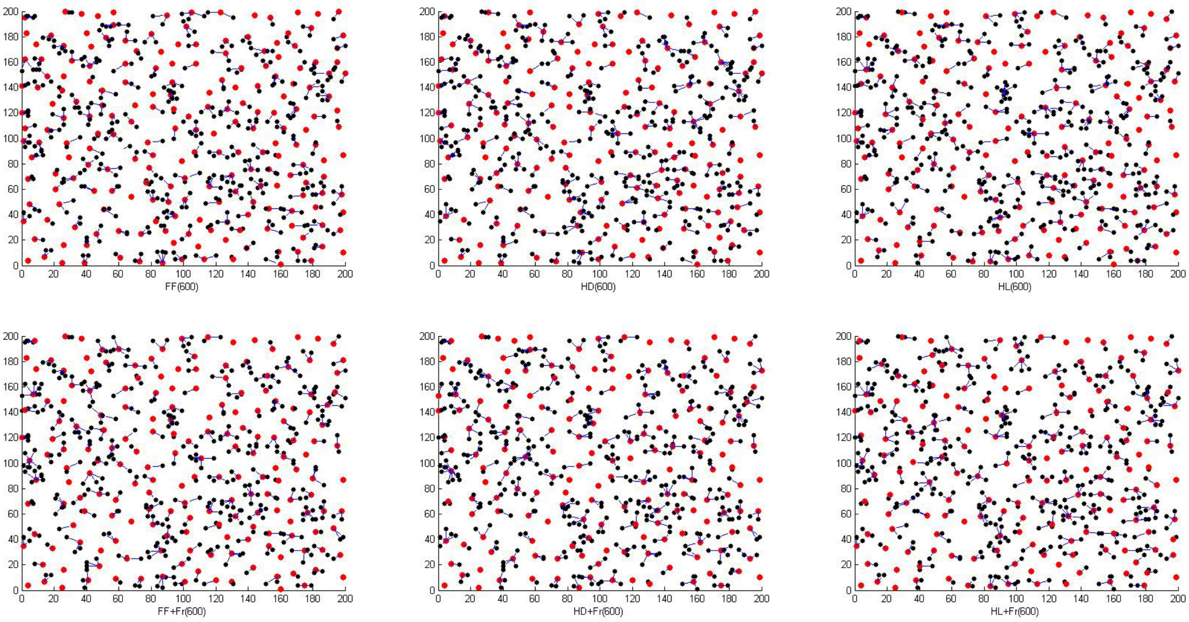

Figure 8. Comparison of 600-node network construction

5 Conclusion

This paper focuses on the study of using as few as possible relay nodes for data gathering and forwarding in wireless sensor networks. If there is only one sink node and all the other nodes are conventional sensors, the problem of energy holes will happen, and the sensors near the sink will exhaust their energy firstly. Constructing one more layer in the network topology is a good strategy to balance the workload of the sensors. Therefore, in this paper FF+Fr(α), HD+Fr(α) and HL+Fr(α) methods are proposed to find the appropriate locations of the relay nodes for building a three-layer network and improve the performances of the farthest-first traversal and nearest-first methods, also enabling them to solve other complicated k-center problems. The simulation results show that FF+Fr(α) can construct a sensor network with 6 stations fewer than the farthest-first traversal method in a 100-node network and with 20 stations fewer in a 600-node network. Moreover, the lower data sensing effective ratio, the fewer duplicated packages will be sent by the sensor nodes, and the less energy will be consumed. The simulation results also show that our schemes consume less energy than the other methods and thus prolong the network lifetime.

Acknowledgments

This work is partially supported by the Fujian Provincial Natural Science Foundation, China (2017J01730) and the Key Project of Fujian Education Department Funds, China (JA15323), Shenzhen

Innovation and Entrepreneurship Project with the project number: GRCK20160826105935160.

References

[1] M. M. Zanjireh, H. Larijani, A Survey on Centralised and Distributed Clustering Routing Algorithms for WSNs, *IEEE 81st Vehicular Technology Conference* (VTC Spring), Glasgow, UK, 2015, pp. 1-6.

[2] T. W. Sung, C. S. Yang, An Adaptive Joining Mechanism for Improving the Connection Ratio of ZigBee Wireless Sensor Networks, *International Journal of Communication Systems*, Vol. 23, No. 2, pp. 231-251, February, 2010.

[3] A. A. Abbasi, M. Younis, A Survey on Clustering Algorithms for Wireless Sensor Networks, *Computer Communications*, Vol. 30, No. 14-15, pp. 2826-2841, October, 2007.

[4] T.-K. Dao, T.-S. Pan, T.-T. Nguyen, A Compact Articial Bee Colony Optimization for Topology Control Scheme in Wireless Sensor Networks, *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 6, No. 2, pp. 297-310, March, 2015.

[5] T.-T. Nguyen, T.-K. Dao, M.-F. Horng, C.-S. Shieh, An Energy-based Cluster Head Selection Algorithm to Support Long-lifetime in Wireless Sensor Networks, *Journal of Network Intelligence*, Vol. 1, No. 1, pp. 23-37, February, 2016.

[6] T. W. Sung. C. Y. Lee. Y. T. Lu, L. C. Shiu, F. T. Lin, Energy-aware Topology Construction and Maintenance for Energy Harvesting Sensor Networks, *the International Conference on Advanced Intelligent Systems and Informatics*, Cairo, Egypt, 2016, pp. 723-731.

[7] D. J. Rosenkrantz, R. S. Stearns, P. M. Lewis, An Analysis of Several Heuristics for the Traveling Salesman Problem, *SIAM Journal on Computing*, Vol. 6. No. 3, pp. 563-581, 1977.

[8] T. F. Gonzalez, Clustering to Minimize the Maximum Intercluster Distance, *Theoretical Computer Science*, Vol. 38, pp. 293-306, 1985.

[9] S. Chechik, D. Peleg, The Fault-tolerant Capacitated K-Center Problem, *Theoretical Computer Science*, Vol. 566, pp. 12-25, February, 2015.

[10] V. V. Vazirani, *Approximation Algorithms,* Springer Science & Business Media, 2013.

[11] C.-F. Huang, Y.-C. Tseng, A Survey of Solutions for the Coverage Problems in Wireless Sensor Networks, *Journal of Internet Technology*, Vol. 6, No. 1, pp. 1-8, January, 2005.

[12] H. Du, Y. Xu, B. Zhu, An Incremental Version of the k-Center Problem on Boundary of a Convex Polygon, *Journal of Combinatorial Optimization*, Vol. 30, No. 4, pp. 1219-1227, November, 2015.

[13] D. Liang, L. Mei, J. Willson, W. Wang, A Simple Greedy Approximation Algorithm for the Minimum Connected k-Center Problem, *Journal of Combinatorial Optimization*, Vol. 31, No. 4, pp. 1417-1429, May, 2016.

[14] R. Ge, M. Ester, B. J. Gao, Z. Hu, B. Bhattacharya, B. Ben-Moshe, Joint Cluster Analysis of Attribute Data and Relationship Data: The Connected k-center Problem, Algorithms and Applications, *ACM Transactions on Knowledge Discovery from Data*, Vol. 2, No. 2, Article No. 7, July, 2008.

[15] A. E. Feldmann, Fixed Parameter Approximations for k-Center Problems in Low Highway Dimension Graphs, *42nd International Colloquium on Automata, Languages, and Programming, Kyoto, Japan*, 2015, pp. 588-600.

[16] O. A. Ghiasvand, M. A. Sharbafi, Using Earliest Deadline First Algorithms for Coalition Formation in Dynamic Time-critical Environment, *International Journal of Information and Education Technology*, Vol. 1, No. 2, pp. 120-125, June, 2011.

[17] P. Mittal, C. Tripathi, A. Kumar, A Centralized Approach for Resolving Physical Interference between Robots using Nearest First Swarm Method, *IEEE 6th International Conference on Computer Science and Information Technology (CSIT)*, Amman, Jordan, 2014, pp. 117-121.

[18] S. Basu, A. Banerjee, R. J. Mooney, Active Semi-Supervision for Pairwise Constrained Clustering, *the 2004 SIAM International Conference on Data Mining*, Orlando, Florida, 2004, pp. 333-344.

[19] D. S. Hochbaum, D. B. Shmoys, A Best Possible Heuristic for the k-Center Problem, *Mathematics of Operations Research*, Vol. 10, No. 2, pp. 180-184, May, 1985.

[20] Sharmila, M. Kumar, An Optimized Farthest First Clustering Algorithm, *2013 Nirma University International Conference on Engineering*, Ahmedabad, India, 2013, pp. 1-5.

[21] G. Sehgal, K. Garg, Comparison of Various Clustering Algorithms, *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 3, pp. 3074-3076, May-June, 2014.

[22] N. Sharma, A. Bajpai, R. Litoriya, Comparison the Various Clustering Algorithms of Weka Tools, *International Journal of Emerging Technology and Advanced Engineering*, Vol. 2, No. 5, pp. 73-80, May, 2012.

[23] D. A. Vadeyar, H. K. Yogish, Farthest First Clustering in Links Reorganization, *International Journal of Web & Semantic Technology*, Vol. 5, No. 3, pp. 17-24, July, 2014.

[24] A. Dharmarajan, T. Velmurugan, Lung Cancer Data Analysis by k-Means and Farthest First Clustering Algorithms, *Indian Journal of Science and Technology*, Vol. 8, No. 15, pp. 1-8, July, 2015.

[25] M. Marta, M. Cardei, Improved Sensor Network Lifetime with Multiple Mobile Sinks, *Pervasive and Mobile Computing*, Vol. 5, No. 5, pp. 542-555, October, 2009.

[26] D. Harel, Y. Koren, Graph Drawing by High-dimensional Embedding, *GD '02 Revised Papers from the 10th International Symposium on Graph Drawing*, Irvine, CA, 2002, pp. 207-219.

[27] A. Baresel, H. Sthamer, M. Schmidt, Fitness Function Design To Improve Evolutionary Structural Testing, *GECCO'02 Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation*, New York, NY, 2002, pp. 1329-1336.

Biographies

Jeng-Shyang Pan received the Ph.D. degree in Electrical Engineering from the University of Edinburgh, U.K. in 1996. Currently, he is a Professor in Harbin Institute of Technology Shenzhen and Fujian University of Technology. His current research interests include soft computing, information security, big data mining and signal processing.

Lingping Kong is a Ph.D. student in Harbin Institute of Technology Shenzhen Graduate School. Her current research interests include wireless sensor network, artificial intelligence etc.

Tien-Wen Sung received his Ph.D. degree in Computer and Communication Engineering from Department of Electrical Engineering at Cheng Kung University, Taiwan in 2014. He is currently an Associate Professor in College of Information Science and Engineering, Fujian University of Technology, China. His research interests include wireless sensor networks and e-learning.

Pei-Wei Tsai received his Ph.D. in Electronic Engineering in 2012. He is the Executive Editor of Journal of Network Intelligence. He is a research fellow at the Swinburne University of Technology in Australia. His research interests include swarm intelligence, information security, big data analysis, intelligent transportation, and wireless sensor network.

Václav Snášel is a researcher and university professor at the VSB-Technical University of Ostrava. His research and development experience includes over 25 years in the Industry and Academia. He works in a multi-disciplinary environment involving artificial intelligence, conceptual lattice, information retrieval, semantic web, neural network, web intelligence etc.