NANZAN-TR-2006-04
(revised version of NANZAN-TR-2006-01)

A New Clustering Algorithm Using Facility Location Theory for Wireless Sensor Networks

Takehiro Furuta, Mihiro Sasaki, Fumio Ishizaki,
Atsuo Suzuki, Hajime Miyazawa

March 2007

Technical Report of the Nanzan Academic Society
Mathematical Sciences and Information Engineering

A New Clustering Algorithm Using Facility Location Theory for Wireless Sensor Networks

Takehiro Furuta [a,*], Mihiro Sasaki [a], Fumio Ishizaki [a], Atsuo Suzuki [a], Hajime Miyazawa [a]

[a] *Faculty of Mathematical Sciences and Information Engineering, Nanzan University*

27 Seirei, Seto, Aichi, 489-0863, Japan

Abstract

In this paper, we study clustering algorithms for wireless sensor networks from a view of *facility location theory*. From this view, we can consider that LEACH-C, which is one of the principal studies on cluster-based network organization, formulates the clustering problem as a p-median problem. We point out drawbacks of the formulation in LEACH-C. To overcome the drawbacks, we formulate the problem as an uncapacitated facility location problem. Computational experiments show that the proposed algorithm can extend the lifetime of sensor networks, compared to LEACH-C.

Keywords: Location, p-median problem, UFLP, Wireless sensor network

1 Introduction

Wireless sensor networks have been paid much attention due to their rich applications in the scientific, medical, commercial and military domain. A wireless sensor network is formed by tens to thousands sensor nodes randomly deployed in a target field. Sensor nodes should be first organized into an ad hoc network and send information about monitored events to a data sink or a remote base station (BS) through the organized network.

[*]Corresponding author. Tel.: +81-561-89-2000, fax: +81-561-89-2083, *E-mail address: takef@fw.ipsj.or.jp*

One of crucial challenges in the organization of sensor networks is energy efficiency, because battery capacities of sensor nodes are severely limited and replacing the batteries is not practical. Once all sensor nodes run out their batteries, the sensor network doesn't work anymore. So various network architectures and protocols to save energy consumption and extend the lifetime of sensor networks have been studied (e.g., see [1] and references therein). Among them, cluster-based network organization is considered as the most favorable approach in terms of energy efficiency. In this approach, sensor nodes are organized into clusters, and one sensor node in each cluster is selected as cluster head (CH) to play a special role as transfer point (see Figure 1). Moreover, each CH creates a schedule for the sensor nodes within the cluster, which allows the radio components of each non-CH-node to be turned off all times except during its transmit time.

The rotation of CHs is also important factor to organize sensor networks. Since the BS is generally far away from the sensor field, CHs exhaust much amount of energy for the data transmission to the BS. Hence, CHs will die quickly if the same node continuously works as a CH. Thus, in order not to drain the battery power of a single sensor, clustering algorithms studied so far introduce the rotation of CHs among sensor nodes.

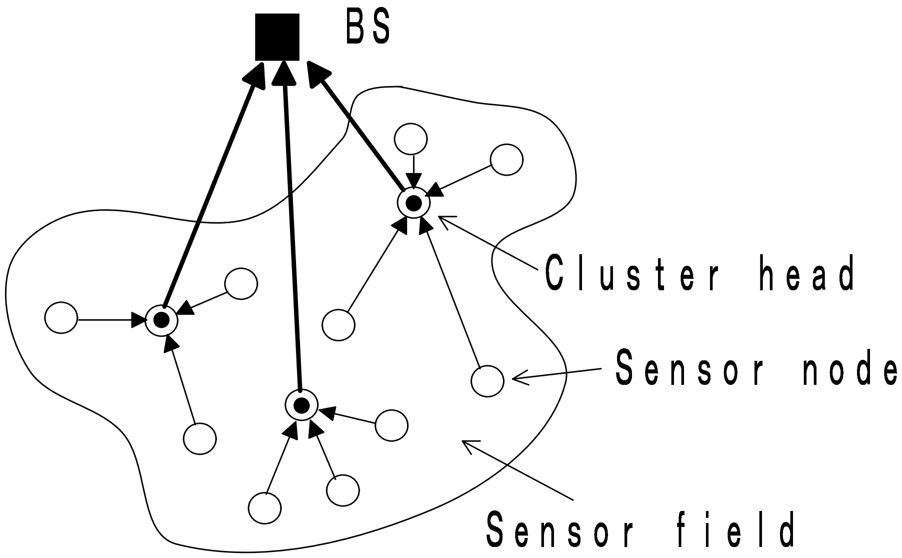

Figure 1: Cluster-based sensor network

One of the primal studies on cluster-based network organization is LEACH (Low Energy Adaptive

Clustering Hierarchy) [2] and LEACH-C (LEACH-Centralized) [3]. LEACH is based on self-organized network. On the other hand, LEACH-C is a centralized cluster formation version of LEACH, where the BS organizes and controls the network. More precisely, LEACH-C protocol provides a centralized cluster formation, local processing for aggregation of sensing data, and the rotation of CHs for every round. These activities are aimed at achieving uniform energy consumption among sensor nodes and maximizing network lifetime. Since the BS usually does not have energy constraint, centralized cluster formation methods can be attractive alternatives. In fact, LEACH-C is more efficient than LEACH in terms of energy consumption from the computational results [3].

In this paper, we study centralized clustering algorithms for wireless sensor networks from a view of *facility location theory*. From this view, we can consider that LEACH-C formulates the clustering problem as a p-median problem [4], which is one of well-known facility location problems. We point out drawbacks of the formulation in LEACH-C. To overcome the drawbacks of LEACH-C, we formulate the clustering problem as an uncapacitated facility location problem (UFLP) [4] incorporating additional factors that haven't been taken into account in LEACH-C.

The remainder of this paper is organized as follows. In Section 2, we point out drawbacks in the clustering algorithm of LEACH-C, and give basic idea to overcome the drawbacks. In Section 3, we formulate the clustering problem as a UFLP. Based on the formulation, we propose a new clustering algorithm for wireless sensor networks. In Section 4, we show computational results to compare the lifetime of sensor network. Finally, we give concluding remarks and mention our future work in Section 5.

2 Centralized cluster-based sensor network

The operation of cluster-based sensor network is usually divided into *rounds*. Each round has two phases, which are the clustering phase and the data transmission phase. Rounds are repeated to monitor events continuously. In the following, we briefly summarize the centralized cluster formation algorithm employed in LEACH-C.

In the clustering phase of LEACH-C, each sensor node first reports information about its current location (this information may be obtained by a GPS receiver) and its battery level to the BS. The BS computes the average battery level of nodes and selects sensor nodes as CH candidates which have above-average battery level. Finally, the BS determines CHs and finds a cluster formation by solving a p-median problem where the objective is to minimize the sum of the squared distances from each sensor node to the nearest CH.

In the data transmission phase of LEACH-C, each non-CH node sends monitored data to its CH. After each CH receives all data from the sensor nodes within the cluster, *data aggregation* process can be carried out by each CH. Since the data monitored by each sensor node within a cluster are often correlated or redundant, the BS does not require all data. Thus, the data aggregation process removes such redundant data and reduces the size of data to send the BS, which consequently saves the energy consumption of CHs.

Heinzelman et al. [3] reported that LEACH-C performs better compared to LEACH from their simulation. However, there are drawbacks listed below in the clustering algorithm of LEACH-C.

- Although it is important to consider energy consumption or remaining battery level of sensor nodes, the algorithm does not take them into account. It simply takes into account the squared distances from sensor nodes to the nearest CHs. Since the distances from CHs to the BS is usually much longer than the distances between sensor nodes, these distances are critical factors governing the network lifetime.

- The algorithm considers the energy consumption only caused by data transmission, although CHs also expend their battery powers on receiving and aggregating data.

- In the algorithm, the number of CHs is predetermined and fixed through rounds. The number of CH candidates may be less than the predetermined number after some rounds. Hence, this can cause the p-median problem to be infeasible due to the shortage of CH candidates, although many sensor nodes are still alive. When it happens, we can not continue rounds with this algorithm unless we select CHs in alternative way.

To overcome the above drawbacks, we formulate the clustering problem as a UFLP where the objective is to maximize the total battery level of all sensor nodes taking into account all kinds of energy consumption. The facilities correspond to the CHs and the fixed cost for each sensor node corresponds to the energy consumption to transmit data to the BS as a CH. By formulating the problem as a UFLP, the number of selecting CHs in each round will be more flexible than using the p-median problem, which finally may bring a long lifetime network.

3 Formulation

As mentioned in the previous section, each round has two phases. The first phase is the clustering phase, where the clustering problem is solved. In this section, we formulate the clustering problem as a UFLP, and also show the p-median problem formulation used in LEACH-C. First, we employ the following notations:

N**:** the index set of sensor nodes,

d_{ij}**:** the distance from sensor node $i \in N$ to sensor node $j \in N$,

f_i**:** the distance from sensor node $i \in N$ to the BS,

b_i**:** the battery level of sensor node $i \in N$ (J),

l**:** the data amount sent by each sensor node (bit),

E**:** coefficient for the radio dissipate (J/bit),

E_{DA}**:** coefficient for data aggregation (J/bit),

n**:** the number of sensor nodes which have positive battery level,

α**:** parameter to determine CH candidates ($0 < \alpha \leq 1$),

S_i**:** 0 if sensor node i has a positive battery level and 1 otherwise.

We assume that every sensor node sends a fixed length of message (l bits) in each round. The parameter α is introduced to allow more flexible CH candidate selection. Note that $\alpha = 1$ throughout in LEACH-C. As mentioned in Section 1, CHs play not only transfer points but also other roles to control sensor networks. However, we omit these factors and assume that CHs play only transfer points.

We assume that the model of the energy consumption of transmitting and receiving the data is the same as LEACH-C [3]. Amplifier energy used for data transmission is defined by two models depending on the distance between the sensor nodes. If the distance is less than the distance threshold d_0, we use the free space model, otherwise, we use the multi-path model. Amplifier energy used for data transmission from sensor node i to j is given by

$$D_{ij} = \begin{cases} \epsilon_{fs} {d_{ij}}^2 & (\text{if } d_{ij} < d_0) \\ \epsilon_{mp} {d_{ij}}^4 & (\text{if } d_{ij} \geq d_0), \end{cases}$$

and that from sensor node i to the BS is given by

$$F_i = \begin{cases} \epsilon_{fs} {f_i}^2 & (\text{if } f_i < d_0) \\ \epsilon_{mp} {f_i}^4 & (\text{if } f_i \geq d_0), \end{cases}$$

where ϵ_{fs} and ϵ_{mp} are coefficients for the two models, respectively. Amplifier energy used for data receiving from a sensor node is lE. For data aggregation, we adopt a perfect data aggregation as LEACH-C, where the received messages is aggregated into one message at CHs. As a result, every CH sends l bits data to the BS.

We further introduce the following decision variables.

x_i: binary variable such that $x_i = 1$ if sensor node $i \in N$ is selected as a CH, and $x_i = 0$ otherwise.

y_{ij}: binary variable such that $y_{ij} = 1$ if sensor node $i \in N$ belongs to the cluster where sensor node $j \in N$ is a CH, and $y_{ij} = 0$ otherwise.

We now propose a new formulation for clustering problem of sensor networks. The clustering problem is formulated as the following integer programming problem:

$$\max \sum_{i \in N} \left\{ b_i - \left(lE + l \sum_{j \in N} D_{ij} y_{ij} + l F_i x_i \right) - lE \sum_{j \in N} y_{ji} - lE_{DA} \sum_{j \in N} y_{ji} \right\} \tag{1}$$

$$
\begin{aligned}
\text{s.t.} \quad & x_i + \sum_{j \in N} y_{ij} + S_i = 1, && i \in N, && (2)\\
& \left(b_i - \frac{\alpha}{n} \sum_{i \in N} b_i \right) x_i \geq 0, && i \in N, && (3)\\
& y_{ij} \leq x_j, && i, j \in N, && (4)\\
& x_i \in \{0, 1\}, && i \in N, && (5)\\
& y_{ij} \in \{0, 1\}, && i, j \in N. && (6)
\end{aligned}
$$

The objective is to maximize the total amount of remaining batteries of sensor nodes after one round. From constraint (2), each sensor node plays a CH or sends the data to the nearest CH as far as its battery level is positive. Constraint (3) ensures that sensor node which has at least α times as much as the average battery level of all alive sensor nodes can be a candidate of CH. Constraint (4) means that only CHs can receive the data.

Note that the objective (1) is to maximize the total sum of battery level of sensor nodes and it can be rewritten as the standard form of the objective in the UFLP:

$$
\min \sum_{i \in N} \sum_{j \in N} (E + D_{ij} + E_{DA}) y_{ij} + \sum_{i \in N} F_i x_i.
$$

Hence, the problem can be regarded as a UFLP.

In a similar manner, we can formulate the clustering problem in LEACH-C as the following p-median problem:

$$
\begin{aligned}
\min \quad & \sum_{i \in N} \sum_{j \in N} (d_{ij})^2 y_{ij}\\
\text{s.t.} \quad & \sum_{j \in N} x_j = p,\\
& (2), (3), (4), (5).
\end{aligned}
$$

Note that in LEACH-C, the objective is to minimize the total sum of squared distances between the sensor nodes and the nearest CHs, and the energy consumption is not directly considered. Also, the parameter α is fixed and equal to 1 in constraint (3), which means that sensor nodes that have battery over the average of all alive sensor nodes can be CH candidates.

4 Computational Experiments

We made computational experiments to compare the performance of our clustering algorithm with that of LEACH-C. We used exact solutions of the clustering problem in the experiments, while Heinzelman et al. [3] used heuristic solutions brought by the simulated annealing algorithm. Our objective is to examine how long we can extend the network lifetime by using our UFLP based formulation for the clustering problem. From this point of view, we should use the exact solutions rather than the heuristic solutions for our computational experiments. We used an optimization software Xpress-MP (2005B) to obtain the exact solutions. All experiments were run on a PC with Intel Pentium 4 processor (2.53GHz) and 512MB RAM.

We used the following physical constants and parameters in the experiments: $b_i = 0.5$ J, $E = 50$ nJ/bit, $\epsilon_{fs} = 10$ pJ/bit/m^2, $\epsilon_{mp} = 0.0013$ pJ/bit/m^4, $E_{DA} = 5$ nJ/bit, $d_0 = 87$ m, $l = 4200$ bit, and various values of α from 0.1 to 1.0. In addition, we assume that $p = 5$ to solve the p-median problems as LEACH-C [3]. We used two types of data sets, where 100 sensor nodes are randomly deployed in a 100 meters square (data1, $\cdots$, data5) and deployed in a 400 meters square (data6, $\cdots$, data10). For convenience, we define the lower left of the squares as ($x = 0$, $y = 0$), and the upper right of the 100 meters square as ($x = 100$, $y = 100$) and the 400 meters square as ($x = 400$, $y = 400$). We assume that the BS is located at ($x = 50$, $y = 175$) in the data sets of the 100 meters square and at ($x = 200$, $y = 475$) in the data sets of the 400 meters square. The average computational time of our clustering algorithm using Xpress-MP is 4.76 seconds per round.

We simulated data transmission from every node to the BS until all sensor nodes died. To evaluate the performance of the clustering algorithms, we introduce the survival rate, which is defined as the percentage of alive sensor nodes over all sensor nodes. Table 1 shows comparisons of the numbers of rounds between our UFLP based formulation with $\alpha = 1.0$ and the p-median based formulation in LEACH-C for the data sets of the 100 meters square. Table 2 also shows comparisons for the data sets of the 400 meters square. Each table shows the numbers of rounds operated until survival rates decrease to 99%, 90%, 70%, 50%, 30%, 10%, and 0%. We use 100 sensor nodes data in the experiments, then 99%

Table 1: Comparisons of the number of rounds for each survival rate (s.r.) between our UFLP-based formulation with $\alpha = 1.0$ (UFLP) and the p-median based formulation in LEACH-C (p-med.) in a 100 meters square

	data1		data2		data3		data4		data5	
s.r.	p-med.	UFLP	p-med.	UFLP	p-med.	UFLP	p-med.	UFLP	p-med.	UFLP
99%	902	902	611(100)	902	907	916	908	920	888	902
90%	907	922	N/A	919	923	933	919	940	896(93)	915
70%	906(90)	933	N/A	933	924(88)	950	920(89)	950	N/A	933
50%	N/A	939	N/A	939	N/A	957	N/A	959	N/A	943
30%	N/A	944	N/A	944	N/A	961	N/A	965	N/A	948
10%	N/A	952	N/A	954	N/A	969	N/A	972	N/A	957
0%	N/A	963	N/A	960	N/A	981	N/A	979	N/A	963

Table 2: Comparisons of the number of rounds for each survival rate (s.r.) between our UFLP-based formulation with $\alpha = 1.0$ (UFLP) and the p-median based formulation in LEACH-C (p-med.) in a 400 meters square

	data6		data7		data8		data9		data10	
s.r.	p-med.	UFLP	p-med.	UFLP	p-med.	UFLP	p-med.	UFLP	p-med.	UFLP
99%	38	49	31	47	36	53	50	59	32	42
90%	53	70	61	74	50	68	76	81	62	74
70%	105	119	102(78)	135	131(78)	171	103	113	90(76)	109
50%	235	245	N/A	180	N/A	247	179	201	N/A	151
30%	254(44)	359	N/A	291	N/A	342	183(50)	392	N/A	271
10%	N/A	441	N/A	423	N/A	463	N/A	545	N/A	502
0%	N/A	526	N/A	455	N/A	469	N/A	550	N/A	575

survival rate means that the first sensor node died. In other words, the numbers appeared in the row labeled “99%” show the number of rounds operated with all sensor nodes alive. Note that the numbers appeared in the row labeled “0%” express the network lifetime. The numbers in parenthesis appeared in the “p-med.” columns are the number of nodes still alive when the clustering problem becomes infeasible. For instance, 906(90) means that the problem becomes infeasible at the 906-th round and ninety of sensor nodes are still alive.

The p-median problem in LEACH-C becomes infeasible due to the lack of CH candidates. Tables 1 and 2 show that many sensor nodes are still alive when the problem becomes infeasible. Especially in the experiment for data2, all sensor nodes are still alive when the problem becomes infeasible at the 611-th round. In other words, only less than five sensor nodes have above-average energy among 100 sensor nodes. This seems unlikely to happen. Let us make sure what happens at the 611-th round.

Figure 2 shows the battery level of each sensor node at the 611-th round of data2. The horizontal line expresses the average remaining battery level. The circled crosses mean that the battery levels are above the average. We can see that only four sensor nodes can be CH candidates though the p-median problem needs at least five candidates to select five CHs. We also find that three of the four candidates have much larger remaining batteries than the others’. From further investigation, we have found that the three candidates are closed to the BS. Thus, they do not expend a large amount of energy to send the data to the BS even if they are selected as CHs. However, in spite of the advantage of their locations being selected as CHs, they are selected as CHs only about the same times as the other sensor nodes. In short, they have enough battery level, but play a role as CHs less frequently. In such a case, only a few sensor nodes may have much larger remaining batteries and raise the average battery level, which possibly makes the rest of the sensor nodes have below-average energy. As a result, this leads to the shortage of CH candidates and makes the problem infeasible at early stage of rounds remaining many alive nodes.

As an alternative, we proposed a UFLP based formulation in which the number of CHs dynamically optimized to maximize the total battery level in each round as shown in Figure 3. This keeps the problem being feasible even after a quite large number of rounds, although sensor nodes do not expend

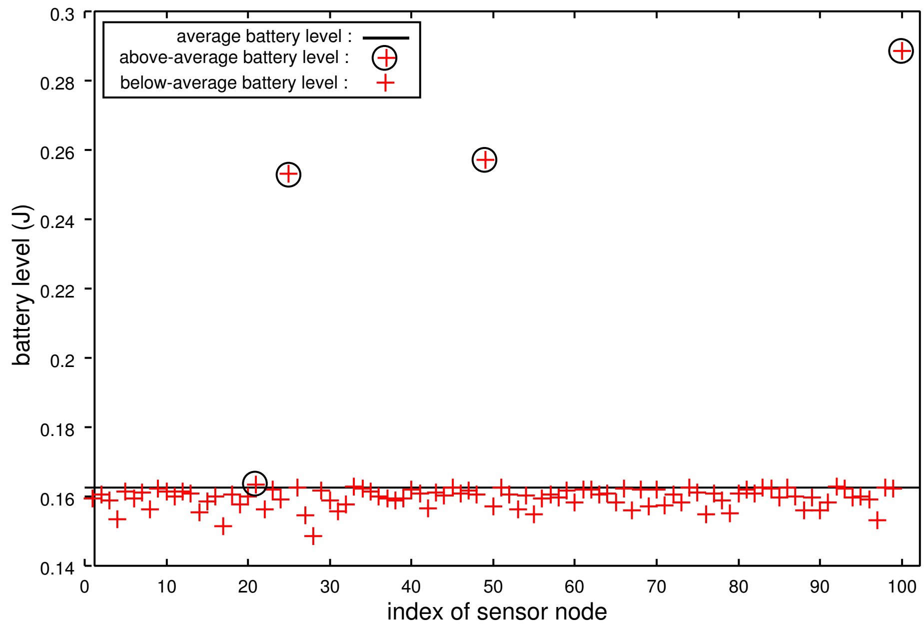

Figure 2: The battery level of each sensor node at the 611-th round of data2 (100×100)

their batteries uniformly.

From Tables 1 and 2, we can observe that, in terms of the network lifetime, our UFLP based formulation is more suitable than the p-median based formulation proposed in LEACH-C. Using our formulation, we can get the exact solutions of the clustering problems without being infeasible until all sensor nodes died. Moreover, the numbers of rounds using our formulation are greater than or equal to those using the p-median based formulation at any of the survival rates where we made an observation.

To see how the number of rounds changes with various values of α using our formulation, we simulated data transmission from every node to the BS until all sensor nodes died. Tables 3 and 4 show the computational results for the data sets of 100 meters square and for the data sets of 400 meters square, respectively. In Tables 3 and 4, the columns labeled "ave." show the average number of rounds operated until survival rates decrease to 99%, 90%, 70%, 50%, 30%, 10%, and 0%, and the columns labeled "std." show the standard deviation. The average numbers of rounds are averaged over the five data sets in each type.

From Tables 3 and 4, we can see that the number of rounds is sensitive to the value of α. Note that the number of CH candidates in each round increases as the value of α decreases. As the value of α decreases, the network lifetime is extended, but the number of rounds operated down to 70% survival

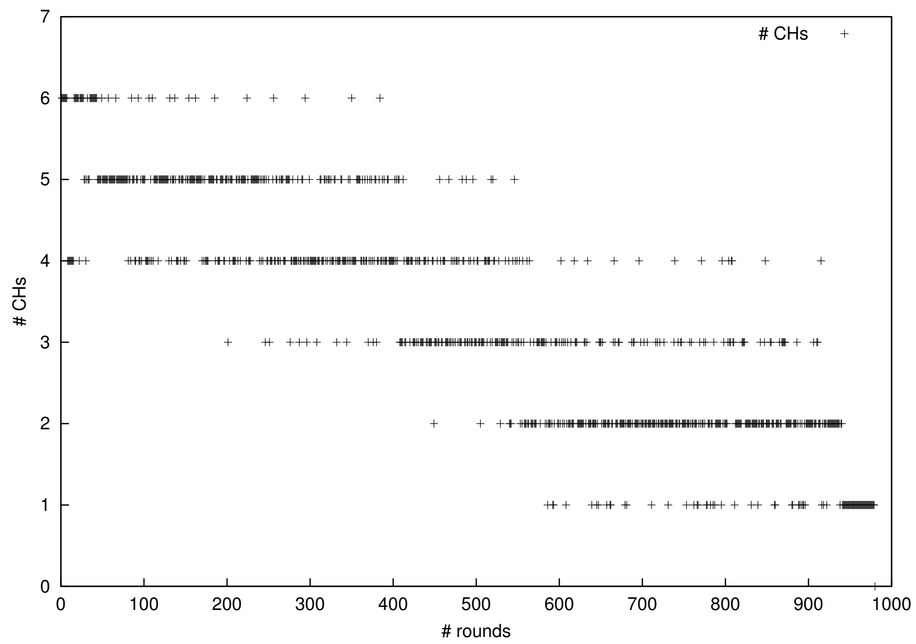

Figure 3: The number of CHs in each round with the case of data1 (100×100) and $\alpha = 0.9$

rate decreases. Thus, when the value of α is small, many sensor nodes die quickly, but the network lifetime tends to be long. On the other hand, when the value of α is large, each sensor node lives longer, but many sensor nodes die suddenly at a time and the network lifetime tends to be short. From these observation, we can select appropriate value of α according to required characteristics of sensor networks. For example, if we need at least 70 % survival rate to monitor events in the sensor field, we should set the value of α rather large. If the network lifetime is more important than the number of alive nodes, we should set the value of α rather small.

5 Conclusion

We formulate the clustering problem as a UFLP and develop a centralized method to find a good clustering with the objective of maximizing the network lifetime. Our method is superior to the clustering algorithm of LEACH-C in many cases. We use 100 sensor nodes examples for computational experiments and get exact solutions using Xpress-MP. Suitable heuristics for the UFLP [5] may enable us to deal with more practical size of sensor network problems.

Table 3: The number of rounds vs. survival rate (s.r.) of nodes : 100×100

	α											
	0.1		0.3		0.5		0.7		0.9		1.0	
s.r.	ave.	std.	ave.	std.	ave.	std.	ave.	std.	ave.	std.	ave.	std.
99%	227.6	17.47	641.6	11.22	903.6	19.27	905.0	12.02	912.8	13.50	908.4	8.88
90%	390.8	25.16	706.0	4.53	922.6	13.94	925.6	14.67	930.4	12.07	925.8	10.38
70%	691.0	23.11	795.4	10.71	933.0	13.78	939.8	12.60	943.8	11.21	939.8	9.31
50%	969.6	14.12	892.6	12.18	943.8	12.21	948.6	13.74	951.2	11.30	947.4	9.84
30%	1222.8	19.41	1095.0	30.65	954.6	11.35	958.6	13.79	957.4	10.71	952.4	9.91
10%	1460.4	28.92	1276.0	29.33	985.8	14.86	977.2	9.68	967.8	11.21	960.8	9.09
0%	1501.0	29.84	1347.4	22.43	1123.6	12.24	1053.6	13.28	988.2	11.14	969.2	9.96

Table 4: The number of rounds vs. survival rate (s.r.) of nodes : 400×400

	α											
	0.1		0.3		0.5		0.7		0.9		1.0	
s.r.	ave.	std.	ave.	std.	ave.	std.	ave.	std	ave.	std.	ave.	std.
99%	20.4	7.70	60.8	14.17	61.0	9.30	57.8	8.90	51.6	7.13	50.0	6.40
90%	67.4	5.77	93.6	10.55	100.8	10.13	95.0	6.40	83.2	6.61	73.4	4.98
70%	118.8	23.15	144.2	29.39	148.4	30.26	145.0	29.01	136.6	28.37	129.4	25.27
50%	176.4	25.73	197.2	39.40	217.2	47.86	218.2	45.89	211.6	39.37	204.8	41.60
30%	278.2	51.06	312.6	52.66	334.4	42.34	342.4	45.40	335.6	51.29	331.0	49.56
10%	495.8	50.48	503.6	43.32	492.6	70.56	489.0	54.36	478.8	51.86	474.8	49.07
0%	813.0	92.43	728.8	92.88	691.8	76.76	615.6	81.86	567.2	79.75	515.0	51.63

References

[1] M. Kochhal, L. Schwiebert, and S. Gupta, "Self-organizing of wireless sensor networks," in *Handbook on Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks* (J. Wu, ed.), pp. 369–392, Auerbach Publications, 2006.

[2] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Hawaii International Conference on System Sciences*, (Maui, Hawaii), January 2000.

[3] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 660 – 670, October 2002.

[4] P. B. Mirchandani and R. L. Francis, *Discrete Location Theory.* Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley & Sons, New York, 1990.

[5] D. Ghosh, "Neighborhood search heuristics for the uncapacitated facility location problem," *European Journal of Operational Research*, vol. 150, pp. 150–162, 2003.