On the Effect of Data Request Message Flooding in Dense Wireless Sensor Networks with a Mobile Sink

On the Effect of Data Request Message Flooding in Dense Wireless Sensor Networks with a Mobile Sink

Daisuke Takaishi *‡, Hiroki Nishiyama*§, Nei Kato*¶, and Ryu Miura†‖
*Graduate School of Information Sciences, Tohoku University, Sendai, Japan
†Wireless Network Research Institute,
National Institute of Information and Communications Technology, Tokyo, Japan
{‡takaishi, §bigtree, ¶kato}@it.ecei.tohoku.ac.jp, ‖ryu@nict.go.jp

***Abstract*—In wireless sensor networks with mobile sink, mobile sink traverse the sensing area and aggregate the data from nodes near the mobile sink. In this scheme, mobile sink can reduce the total energy consumption by dividing the network to the multiple clusters. Since, energy consumption is proportional to the sum of square of communication distance, dividing the network into smaller clusters can reduce the amount of energy required for transmission. Previous researches concluded that increasing number of the cluster reduces energy required for data transmission. However, these ideas do not take into account the energy consumption due to data request message flooding in dense networks. In this paper, we focus on the data request flooding problem which is the massive energy consumption for data request message in dense network. Moreover, we point out that energy consumption of data transmission and data request message are controlled by the number of clusters in the network.**

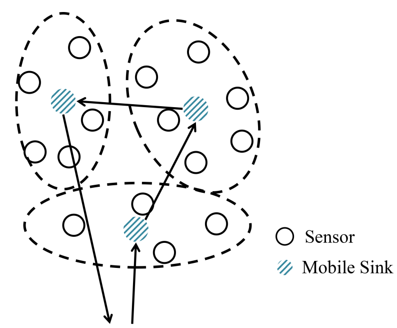


Fig. 1. Data aggregation with a mobile sink

I. Introduction

Wireless Sensor Networks (WSN) have drawn great attention for being a possible candidate for deploying ubiquitous networks. A ubiquitous network refers to a network environment which enables access anytime, anywhere, by any device and by anyone. Recently many applications on ubiquitous networks have been developed such as sharing traffic congestion information by using car navigation systems [1], intruder detection [2] and environment observation [3]. To realize these applications, information such as traffic information, intruder information and temperature information needed to be aggregated to a certain point.

A WSN is consisted of tens of thousand of sensor nodes, of which each is composed of sensing and wireless communication module [4]. In WSN, these nodes communicate with neighboring nodes and form the network automatically. The overall information gathered by the nodes are aggregated to a sink node. Nodes far away from the sink node have to transfer the data in multihop fashion to the sink node. Therefore, the intermediate nodes have to relay data. The nodes, which are close to the sink are likely to relay more data and since relaying data consume energy, the nodes closer to the sink are more likely to run out of energy. Once the battery runs out, the node can no longer sense or transmit data. This problem is called the “energy hole problem” which lead to a short network lifetime [5].

To solve this problem, data aggregation using mobile sink has gained much attention. In WSN with mobile sink, the mobile sink divides the network into some clusters. After clustering, mobile sink traverse each cluster and aggregate the information generated by the nodes that belong to the cluster. Because energy consumption is proportional to the sum of square of communication distance, the use of mobile sink can reduce total energy consumption by shortening the communication path. To reduce energy consumption, the trajectory of the sink node is decided by sensor nodes’ location. Previous researches concluded that increasing number of cluster reduces energy consumption for data transmission. In addition, the upper limitation of number of cluster is decided by some extra parameters (i.e., buffer size and delay). This is because increasing the number of cluster means smaller cluster size and shorter data transmission path length. However, these researches do not take into account energy consumption caused by data request messages. Data request messages have to be sent from the mobile sink to invoke nodes to send data. Nodes which received data request message have to rebroadcast the data request message to their neighboring nodes regardless of their cluster affiliation, since such information is not available for each node. Therefore increasing number of cluster means the number of data request message increases. This problem dramatically degrade the performance in the WSN where sensor are deployed densely.

In this paper, we first present the energy minimized clustering algorithm by using the Expectation-Maximization (EM) algorithm for 2-dimensional Gaussian mixture distribution [6]. The adapted EM algorithm aims to minimize the sum of square of wireless communication distance, because energy consumption is proportional to square of communication distance. Moreover, we first focus on the “data request flooding problem” to point out that too many clusters results in performance degradation.

The reminder of this paper is organized as follows. Section II describes the related works. Section III presents proposed data aggregation scheme. The data request flooding problem is described Section IV. The performance of our method is evaluated in Section V. Finally, Section VI conclude the paper.

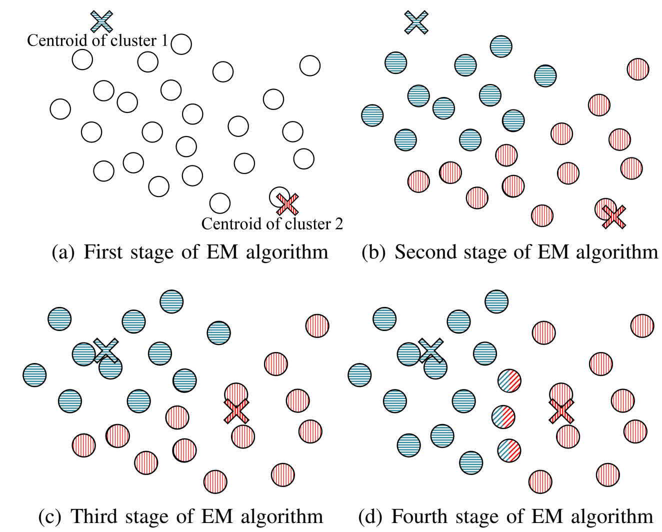
II. Related Work

In recent literature, many studies have already focused on data aggregation with mobile sink(s) in WSNs. Shah et al. proposed the data aggregation scheme with mobile sink having random walk mobility [7]. This scheme is called "Data MULEs," and it aims to reduce power consumption. In the data aggregation scheme, the mobile sink node divides nodes into grids regardless of nodes' location, and patrols grids by using random walk between neighboring grids. This type of clustering which is independent to nodes' location may result in inefficient data aggregation. If there is no node in the cluster, patrolling the cluster is only a waste of time. Determination of suitable routes for mobile sink are critical to reducing energy consumption and prolonging network lifetime.

In contrast to the random clustering algorithm, methods of minimizing energy consumption for data transmission by non-random clustering have been studied. Low-Energy Adaptive Clustering Hierarchy (LEACH) [8] is one of the most famous clustering algorithms. In LEACH, clustering algorithm is executed by each sensor node. Nodes exchange their own remaining energy information, and node which has higher remaining energy is given higher probability of being a cluster head. Energy consumption of the each node becomes probabilistic equal. However, LEACH is based on the assumption that each node can communicate with every other nodes. In the real network, this assumption will result in higher energy consumption. The k-CONID [9]algorithm is a probabilistic algorithm. Nodes exchange its random ID each other and the node which has minimum ID within k-hop is selected as a cluster head.

Because these distributed clustering algorithms based on information exchange, the algorithm considers a limitation of communication range. KAT mobility [10] is a cluster based data aggregation scheme, which uses the K-means algorithm for clustering, Traveling Salesman Problem (TSP) for minimizing traveling time, and Directed Diffusion [11] for data aggregation. KAT mobility aims at not only reducing energy consumption but also at increasing the efficiency, which is the ratio of the aggregated data volume to the consumed energy. KAT mobility creates clusters based on the positions of the nodes in the considered WSN. Therefore, when some nodes become inactive due to low battery, KAT mobility can re-evaluate the trajectory of the mobile sink node and select a more suitable path for performing data aggregation.

Previous researches [12]–[15] consider that increasing number of cluster reduces energy consumption for data transmission because increasing number of cluster means smaller cluster size and shorter data transmission path length. Some researches have considered a certain limitation of number of cluster. In [12] and [13], the limitation is maximum acceptable latency of data aggregation. In [14] and [15] authors define the limitation by nodes buffer size. These limitations are realistic assumption. However, these ideas do not consider energy consumption caused by data request messages. In this paper, we first focus on the effect of data request message by increasing number of cluster.



(a) First stage of EM algorithm (b) Second stage of EM algorithm

(c) Third stage of EM algorithm (d) Fourth stage of EM algorithm

Fig. 2. Overview of EM algorithm

III. Data Aggregation with a Mobile Sink

In this section, we explain simple clustering algorithm which uses nodes' location to minimize the total energy consumption. In this clustering algorithm, we utilize EM algorithm over the 2-dimentional Gaussian mixture distributions since EM algorithm can minimize the sum of square of distance between each node and the centroid of cluster. After clustering, mobile sink traverse the centroids of the cluster computed by EM algorithm. Traversal path is decided by Traveling Salesman Problem (TSP) algorithm. At the centroids of the cluster, mobile sink broadcast the data request message and collect data.

A. EM-based Clustering Algorithm

In order to minimize the energy consumption in WSNs, transmission distance is one of the most important parameters because the required energy for wireless transmission is proportional to the square of the transmission distance. The EM algorithm includes minimizing the sum of the squares of distances between the nodes and the cluster centroid. Moreover, the EM algorithm groups the nodes into a certain number of clusters to reduce energy consumption. The two-dimensional EM algorithm is only based on an assumption that nodes are distributed according to a 2-dimentional Gaussian mixture distribution.

Fig. 2 shows overview of the EM algorithm. Fig. 2(a) is the initial status of network. Every nodes do not belong to any cluster, and centroids of clusters which represented by the cross are randomly decided. At the first step, shown in fig.2(b), EM algorithm calculates each node's degree of dependence that is referred to as responsibility. The responsibility value shows how much a node depends on a cluster. This responsibility value is calculated by the nodes location and centroid location of cluster. Normally, each node depends only on one cluster. However, it is possible for nodes to depend on more than one clusters so that those nodes will not focus their energy in a single cluster. After EM algorithm calculate the responsibility value, it calculate the centroid of the cluster by using node's location as shown in fig. 2(c). The centroid of each cluster is calculated to minimize the distance between each nodes which belong to the cluster and the centroid. After the calculation of centroids, EM algorithm recalculates each node's responsibility value because the centroids location changed in the previous step.

EM algorithm can minimize the sum of squares of the distances between nodes and cluster centroids, however, energy consumption is proportional to the sum of square of communication distance. Communication distance is the sum of distance between each hop from the communicating node to the cluster centroid. Therefore, we need to adapt the EM algorithm to minimize the sum of communication distance. The difference of communication distance and the direct distance between every node and centroids becomes shorter when node density increases. Therefore, we adapt to the EM algorithm by using Gausian mixture distribution and nodes' density.

B. The trajectory of the mobile sink

After clustering of WSN nodes, we will determine the actual trajectory of the mobile sink. The mobile sink traverses through clusters and aggregates data from various nodes. Since it is possible to increase efficiency by reducing the traveling time, it is preferable that the mobile sink traces the shortest path among the cluster centroids. Therefore, we use the solution of TSP. Since TSP is an NP-hard problem, we resort to an approximate solution of TSP as the trajectory.

C. Mobile Sink Data Collection

After arriving at the centroid, the mobile sink will broadcast data request message to nodes, which will in turn try to send the data collected to the mobile sink either directly or via other nodes in a multi-hop fashion. In addition to sending the collected data, the nodes will also rebroadcast the data request message to their neighbor nodes.

IV. Data Request Flooding Problem

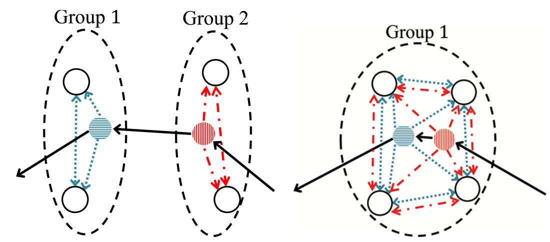
The data aggregation scheme presented in the previous section aims to minimize energy consumption for data collection. However it still has remaining issue in determining the appropriate number of clusters. The previous researches mentioned in the previous section consider that increasing number of cluster reduces energy consumption for data transmission. However, these researches do not take into account the energy consumption for data transmission. Here, we point out that optimal number of clusters can be derived in terms of power consumption due to data request flooding problem, which is affected by the network connectivity.

A. Definition of connectivity

To analyze correlation between energy consumption and connectivity, we need to formulate connectivity of nodes. In the previous literatures, connectivity, C, is often defined by the following equation

$$C = \frac{\sum_{g=1}^{G} N_g(N_g - 1)}{N(N-1)}, \tag{1}$$

where G is a total number of groups and N is a total number of nodes. And N_g is a number of nodes which belong to group g. In this paper, we define a group to be a set of nodes which can communicate with each other directly or in multihop. This metric takes a value between 0 and 1. When all nodes can communicate with each other, value of connectivity is 1. If all node are isolated, the value is 0.



(a) Low connectivity network (b) High connectivity network

Fig. 3. Data request flooding problem

B. Data request flooding problem

In WSNs with mobile sinks, sink node which arrived at the centroid of the cluster, have to send data request message to invoke data transmission from sensor nodes. In general, data request message is transmitted with maximum transmission range without considering neighboring nodes' location and affiliation because information on the nodes' location and cluster information are not available at node without employing additional specific functionality. The nodes which received data request message send the data to the sink nodes and rebroadcast data request message their the neighboring nodes. This data request message is repeatedly rebroadcasted until all nodes have received the message. This becomes a problem, because the network will be flood with redundant data request message causing high energy consumption. Thus, reducing data request transmission is also important for mobile sink scheme.

The impact of data request flooding issues becomes significant when nodes' density becomes larger as shown in Fig.3. There are four sensor nodes and the sink nodes traverse the two points, and broadcast the data request message. In the case of Fig.3(a) where nodes can only communicate with the nodes, which belongs to the same group. Sink node broadcasts data request message to each node in group 1, and because these nodes rebroadcast the data request message, these nodes receive data request message twice. Therefore, the sum of the transmission of data request message of both group 1 and group 2 is 4. On the other hand, Fig.3(b) shows a situation where nodes can communicate with all nodes because all nodes belongs to same group, the data request message is transferred to all nodes. Furthermore, all nodes rebroadcast the data request message. Therefore, sum of the transmission of data request message within group 1 is 8.

Even if the number of nodes and cluster stay the same, the problem become more serious with higher connectivity. Moreover, it is clearly understood that the total number of transmitted data request messages increase when number of cluster is increased. Because of this problem, we need to carefully choose the number of cluster based on connectivity.

C. Effect of connectivity for data request transmission

Consider a group of nodes which has N_g nodes and K_g centroids of cluster. Data request message is sent from every cluster and every nodes rebroadcast it once. Therefore, total required energy to transmit data request message is formulated as follow

$$E_{\text{Req}} = \sum_{g=1}^{G} K_g N_g R^2, \tag{2}$$

where R is maximum transmission range of sensor node. This equation is based on simple assumption which is required energy for transmission is proportional to the square of data transmission range. For simplicity, constant variables are omitted. If N and N_g are sufficiently larger than 1, i.e., $N \gg 1$ and $N_g \gg 1$, then the connectivity, C, can be approximated as follow

$$C = \frac{\sum_{g=1}^{G} N_g(N_g - 1)}{N(N-1)} \div \frac{\sum_{g=1}^{G} N_g^2}{N^2}. \tag{3}$$

Therefore, from equations, 2 and 3, E_{req} is transformed into follows.

$$E_{\mathrm{Req}} = KNR^2C. \tag{4}$$

This analysis says that required energy for data request transmission is proportional to connectivity. Moreover, the function is monotonically increasing function of K, which means lower number of cluster is better for reducing energy consumed by data request transmission.

V. Performance Evaluation

We conducted performance evaluation by using a clustering simulator written by C++. In this section, we first evaluate the clustering efficiency of each clustering algorithm. In the second experiment, we evaluate the energy consumption caused by the data request flooding problem.

A. Energy consumption for data transmission

In this experiment, we evaluate energy consumption for data transmission and efficiency. The efficiency value simply shows how much data can be aggregated per unit of energy.

Table.I shows environment of the first experiment. Sensors are randomly deployed in a 5000 × 5000 square meters. The nodes' communication range are set 438.57 meters. We measure E_{Dat} and efficiency of adapted EM algorithm scheme described in previous section by varying the number of nodes. The first one is data transmission energy E_{Dat}represented in following equation.

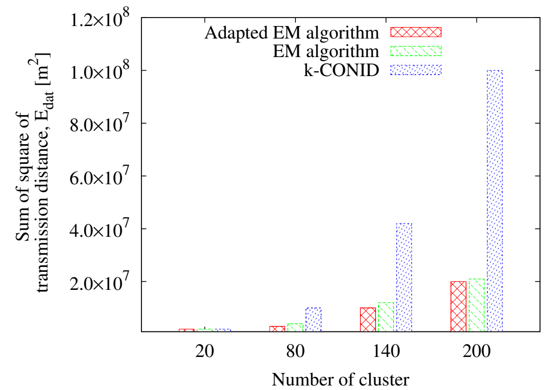
$$E_{\mathrm{Dat}} = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{h=1}^{H_{nk}} \gamma_{nk} \cdot l_h^2, \tag{5}$$

where H_{nk} is number of hop count from nth node to kth centroid and l_h is communication distance of each hop. For simplicity, constant variables are omitted. E_{Dat} shows how well the clustering algorithm works. If locations of every centroids are far away from nodes and cannot connect, E_{Dat} value is calculated as 0. This value does not refer to energy saving, but it is referred to clustering failure. The clustering algorithm which do not consider connectivity sometimes causes this failure (e.g. pure EM algorithm). Thus, we also use metric efficiency and E_{Dat}. The efficiency value is different among clustering algorithm, and if efficiency value is large, the energy consumption for data transmission is also large. We use pure EM algorithm and k-CONID as comparison. EM algorithm is centralized and do not consider connectivity, and k-CONID is distributed and considers connectivity. This distinct characteristics are good for comparison.

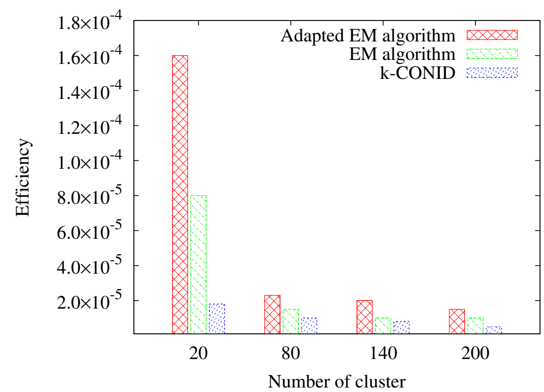
Figure 4(a) and 4(b) are experimental results of required energy and efficiency respectively. Figure 4(a) shows required energy for data transmission is substantially different from k-CONID. The reason for this difference is based on the difference between centralized and distributed nature of the algorithm. The centralized algorithm can calculate a more efficient clustering than the distributed one. Adapted EM algorithm behaves similar to EM algorithm but requires less energy. This improvement comes from the fact that adapted EM considers connectivity and communication distance. Figure 4(b) shows that pure EM algorithm is the worst clustering algorithm when node density is low. Because pure EM algorithm do not consider node connectivity and is a centralized algorithm, clustering results sometimes loses its connectivity to sensor nodes. Thus, when number of nodes is low and node density is small, centroids of EM algorithm can connect only small number of nodes.

TABLE I. Environments of 1st. experiment

Node distribution	Uniformly random
Number of cluster, K	10
Number of node, N	20 - 200
Length of one side of field, L	5000 [m]



(a) Required transmission energy E_{Dat}



(b) Efficiency of each clustering algorithm

Fig. 4. Energy consumption for data transmission and efficiency

B. Data request message flooding problem

In this experiment, we evaluate the effect of data request message flooding problem. Table II shows the environment of second experiment. We evaluate the energy consumption by varying the connectivity, however, connectivity is calculated after scattering nodes. We cannot set connectivity to the value we want. Thus, we use the average results of connectivity.

Fig. 5 shows required energy for data transmission and for

TABLE II. ENVIRONMENT OF 2ND. EXPERIMENT

Number of cluster, K	5 - 50
Number of node, N	50 - 200
Communication range, R	200 - 600 [m]
Length of one side of field, L	1000 - 5000 [m]
Clustering algorithm	Adapted EM algorithm

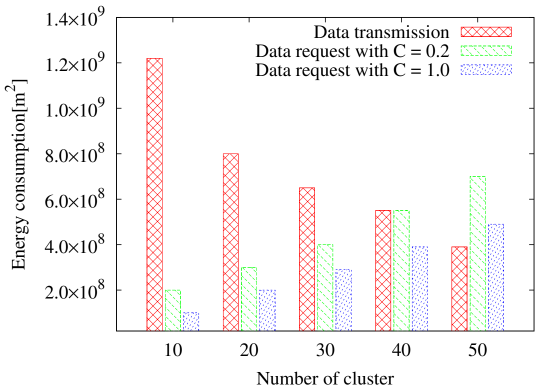


Fig. 5. Required energy for data transmission and data request

data request message in our proposed clustering scheme in the network where connectivity is 0.2 and 1. Connectivity 1 means that each node can communicate with each other directly or in multihop fashion. Therefore, the data request message from the mobile sink is transmitted to every nodes.

This results shows that required energy for data transmission proportionally decreases with increasing number of cluster. However, the energy consumption for data request is proportionally increases with the increasing number of clusters. From these results, to minimize energy consumption for data aggregation, the number of clusters should be decided by considering energy consumption for data transmission and for data request. Moreover, the energy consumption for data request dramatically increases in high connectivity or dense network. This is because in the high connectivity network there is much higher chance of nodes outside the intended cluster to overhear the data request message that would cause the overheard nodes to rebroadcast the message to their neighbors who in turn will rebroadcast the message to their neighbor. This is a waste of energy since the nodes outside the intended cluster have to rebroadcast the data request message even when the message is not intended for them. However, according to the result from the first experiment, the energy consumption of data transmission varies amongst different clustering algorithms. Therefore the optimal number of cluster that should be used is also decided by clustering algorithms.

VI. CONCLUSION

In this paper, we presented adapted EM clustering algorithm. The simulation verify that adapted EM algorithm scheme can increase efficient for different numbers and distributions of nodes. And also simulation shows that our clustering algorithm is the most efficienct among clustering algorithm. In addition, we pointed out that increasing number of cluster increases energy consumption of data request message because of data request flooding problem. The simulation results verify that large number of clusters reduces the energy consumption of data transmission, but increases the energy consumption of data request message. Moreover, energy consumption for data request message increases dramatically in dense, highly connected network because of data request flooding problem. From this result, number of the cluster should be decided by the energy consumption of data transmission and data requst message.

ACKNOWLEDGMENT

Part of this work was conducted under the national project, Research and Development on Cooperative Technologies and Frequency Sharing Between Unmanned Aircraft Systems (UAS) Based Wireless Relay Systems and Terrestrial Networks, supported by the Ministry of Internal Affairs and Communications (MIC), Japan.

REFERENCES

[1] M. Sichitiu and M. Kihl, "Inter-vehicle communication systems: a survey," *Communications Surveys Tutorials, IEEE*, vol. 10, no. 2, pp. 88 –105, quarter 2008.

[2] S. He, J. Chen, Y. Sun, D. Yau, and N. K. Yip, "On optimal information capture by energy-constrained mobile sensors," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 5, pp. 2472–2484, 2010.

[3] W.-Z. Song, R. Huang, M. Xu, B. Shirazi, and R. LaHusen, "Design and deployment of sensor network for real-time high-fidelity volcano monitoring," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 11, pp. 1658 –1674, Nov. 2010.

[4] Y. Z. Xianghui Caoa, Jiming Chena and Y. Suna, "Development of an integrated wireless sensor network micro-environmental monitoring system," *ISA Transactions*, vol. 47, no. 3, pp. 247–255, 2008.

[5] A. Abdulla, H. Nishiyama, J. Yang, N. Ansari, and N. Kato, "Hymn: A novel hybrid multi-hop routing algorithm to improve the longevity of wsns," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 7, pp. 2531 –2541, july 2012.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning*. Singapore: Springer.

[7] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 215 – 233, 2003, sensor Network Protocols and Applications.

[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[9] K. T. Mai and H. Choo, "Connectivity-based clustering scheme for mobile ad hoc networks," in *Research, Innovation and Vision for the Future, 2008. RIVF 2008. IEEE International Conference on*, Jul. 2008, pp. 191 –197.

[10] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, "Fault-resilient sensing in wireless sensor networks," *Computer Communications archive*, vol. 30, no. 11-12, pp. 2375–2384, 2007.

[11] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed diffusion for wireless sensor networking," *IEEE Netw.*, vol. 11, no. 1, pp. 2–16, Feb. 2003.

[12] G. Xing, T. Wang, W. Jia, and M. Li, "Rendezvous design algorithms for wireless sensor networks with a mobile base station," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, ser. MobiHoc '08. New York, NY, USA: ACM, 2008, pp. 231–240.

[13] W. Liang, J. Luo, and X. Xu, "Prolonging network lifetime via a controlled mobile sink in wireless sensor networks," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, dec. 2010, pp. 1 –6.

[14] W. Liu, J. Wang, G. Xing, and L. Huang, "Throughput capacity of mobility-assisted data collection in wireless sensor networks," in *Mobile Adhoc and Sensor Systems, 2009. MASS '09. IEEE 6th International Conference on*, oct. 2009, pp. 70 –79.

[15] G. Keung, B. Li, and Q. Zhang, "Message delivery capacity in delay-constrained mobile sensor networks: Bounds and realization," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 5, pp. 1552 –1559, may 2011.