PERFORMANCE ANALYSIS OF LARGE-SCALE WIRELESS SENSOR NETWORK ARCHITECTURE WITH MULTI-CLUSTER CONFIGURATION

Masashi Sugano
School of Comprehensive Rehabilitation
Osaka Prefecture University
3-7-30, Habikino
Habikino, Osaka 583-8555, Japan
email: sugano@rehab.osakafu-u.ac.jp

Yuichi Kiri and Masayuki Murata
Graduate School of Information Science and Technology
Osaka University
1-5, Yamadaoka
Suita, Osaka 565-0871, Japan
email:{y-kiri, murata}@ist.osaka-u.ac.jp

ABSTRACT

Large-scale sensor networks containing many sensor nodes are often divided into clusters so large that necessary to use multi-hop communication between cluster heads when collecting data. In this paper we derive analytically how parameters such as the communication distance of a cluster head affect the power consumption and throughput of such networks. We also that power consumption can be reduced and throughput roughly doubled by not unnecessarily relaying data in the neighborhood of the sink node.

KEY WORDS

sensor network, simulation, throughput, energy consumption

1 Introduction

Expansion of the application domain of sensor networks has resulted in a demand for large-scale networks of sensor nodes, but the communication ranges of sensor nodes has been restricted to 10-100 m by the need to prolong battery life by reducing energy consumption. To build a large-scale network, we therefore need to use multi-hop communication between sensor nodes. Moreover, from a viewpoint of network control, such as routing, it is easier to implement distributed control over nodes in a hierarchical architecture formed by clustering than it is to control many nodes in a flat architecture. That is, the sensing area should be divided into two or more clusters that each contain a cluster head collecting the data gathered in the sensor nodes and transmitting it to a sink node. Since the load on the sensor node operating as a cluster head becomes large, battery life should be prolonged by changing the cluster heads periodically.

All the data gathered in a sensor network is collected by the sink node. If the cluster head does not perform a data processing, such as data fusion, all the packets generated by the sensor nodes will be sent to the sink node as they are. The situation therefore differs from that in the usual ad hoc network of nodes exchanging data because load concentrates around a sink node. The receiving channel of the sink node thus acts as a bottleneck restricting network throughput. Moreover, the energy consumption of cluster heads close to the sink node becomes large because relay data accumulates there. As a result, network lifetime will also be determined by the load on the sensor nodes in the neighborhood of the sink node. The performance of large-scale sensor networks therefore cannot be predicted without first clarifying the influence of parameters determining the load in the neighborhood of the sink node.

Duarte-Melo and Liu [1] evaluated the performance of flat and hierarchical sensor networks analytically and showed that throughput is improved by clustering. Communication from a cluster head to the sink node is performed in a single hop, however, and the network characteristic due to multi-hop communication between cluster heads is not clear. In this paper we analytically derive the ways in which the throughput and energy consumption of a large-scale sensor network using multi-hop communication between cluster heads are influenced by parameters related to the transmission range of cluster heads. Furthermore, we show how a throughput and the amount of energy consumption can be improved by reducing the number of unnecessary relays in the sink node neighborhood.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 presents our network model. In Section 4 we analyze energy consumption and compare the results of that analysis with the results of simulation experiments. In that section we also show how much energy consumption can be reduced by avoiding unnecessarily relaying data in the neighborhood of the sink node. Section 5 presents the analysis of throughput, and Section 6 concludes the paper.

2 Related work

Various groups of investigators have explored the relationship between the data collection systems used in sensor networks and the power consumption of those networks, and a low-energy adaptive clustering hierarchy (LEACH [2]) has been proposed to reduce the power consumption of sensor networks using single-hop communication within clusters and from cluster heads to the sink node.

PEGASIS (power-efficient gathering in sensor information systems) [3], on the other hand, is a chain-based protocol that minimizes energy consumption by having

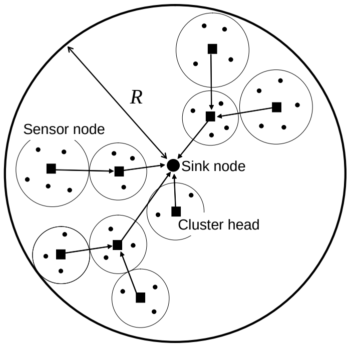


Figure 1. Network model

sensor nodes communicate only with their nearest neighbors. PEGASIS enables CDMA-capable nodes to transmit data simultaneously with little interference and even non-CDMA nodes to transmit data simultaneously if they are spatially separated and linked in a chain-based 3-level hierarchy.

LEACH attracted attention as a protocol letting sensor nodes form clusters autonomously, and many improvements have since been proposed. LEACH assumes single-hop communication within a cluster, but an energy-aware routing protocol taking into account the possibility of multi-hop communication within a cluster has also been proposed [4], as has a two-phase clustering scheme for saving energy at the cost of increasing delay [5].

Research efforts have also been devoted to improving the formation of clusters. Since a cluster head is determined randomly in LEACH, the numbers of sensor nodes in clusters will vary and this variation will reduce the network lifetime. Clustering methods that yield little variation in the number of nodes have therefore been proposed [6]. The hybrid energy-efficient distributed clustering system (HEED) produces a fairly uniform distribution of cluster heads while taking load and residual energy into consideration [7]. The hybrid indirect transmission (HIT) clustering protocol further reduces energy consumption and network delay by having each sensor node compute a TDMA schedule controlling its data transmission [8].

3 Network model

A model of the sensor network under consideration is shown in Fig. 1, and the following notations and assumption are introduced:

- The sensing area is a circle of radius R.
- The total number N of nodes and the number N_{CH} of cluster heads shall be uniformly distributed over the whole sensing area.
- Each node uses the same communication channel when it serves as a sensor node, and it uses another same communication channel when it serves as a cluster head.
- The transmitted electric power of each node shall be controlled such the transmission range is r when that node acts as a sensor node and is r_{CH} when that node acts as a cluster head.
- The data transmitted by the node X_i is received correctly received by node X_j only under the following conditions [9]:
  1. Node X_j is within the transmission range of node X_i. That is,
  $$|X_i - X_j| \leq r.$$
  2. For any other node X_k,
  $$|X_k - X_j| \geq |X_i - X_j| + \Delta.$$
  Here $\Delta > 0$ is a guard zone for avoiding the interference from other.
- The whole sensing area shall be divided into clusters and one cluster head shall be put on the center of each cluster.
- A sensor node transmits data to its cluster head in a multi-hop fashion, and each cluster head transmits its collected data to the cluster head of nearest adjacent cluster closer to the sink node.
- Data processing, such as a data fusion, is not performed. That is, the k bit data generated by a sensor is sent to a sink node as it is.
- The power consumption model of a sensor node is the one used when the LEACH was proposed [2]. That is, a sensor node needs energy E_{elec} (nJ/bit) energy for the operation of its transceiver circuit and needs energy ε_{amp} (pJ/bit/m^2) for amplification of its transmitted signal.

4 Analysis of energy consumption

4.1 Relation between energy consumption and the distance from the sink node

According to the model specified in Section 3, the energy consumed by a sensor node in the target system can be calculated as follows. The following energy is consumed when operating as a sensor node:

- Energy for transmitting the data it sensed or received from another sensor node to the cluster head or the nearest sensor node closer to the cluster head.
- Energy for receiving the data to be relayed from another sensor node within the same cluster but farther from the cluster head.

Moreover, the following energy is consumed when operating as a cluster head:

- Energy for receiving the data from sensor nodes within its cluster.
- Energy for receiving data to be relayed from a cluster head farther from the sink node.
- Energy for transmitting data to the sink node or the nearest cluster head closer to the sink node.

First we derive energy consumption when the node acts as an ordinary sensor node. Although the energy consumed for relaying data will depend on the position of the node within its cluster, for simplicity here we approximate it by using the method of [10] to calculate the energy consumption of the whole cluster and dividing that by the number of the nodes in the cluster.

Here we consider the case in which the k bit generated in all the sensor nodes within a cluster is collected in the cluster head. The number of sensor nodes in each cluster, N_c , is

$$N_c = \frac{N}{N_{CH}}. \tag{1}$$

If all the number of times of transmission within a circular cluster of radius l is x_c, the number $n(h)$ of nodes left for h hops to the cluster head will be calculated as follows:

$$n(h) = N_c \frac{r^2}{l^2} \{h^2 - (h-1)^2\}. \tag{2}$$

Moreover, for one node the number $l(h)$ of times of transmission left for h hops to the cluster head is calculated as follows:

$$l(h) = \frac{\sum_{i=h}^{\lceil \frac{l}{r} \rceil} n(i)}{n(h)} = \frac{N_c\{1-(h-1)^2 \frac{r^2}{l^2}\}}{n(h)}. \tag{3}$$

Therefore, the total number of times of transmission in a cluster, x_c, is

$$x_c = \sum_{h=1}^{\lceil \frac{l}{r} \rceil} n(h) l(h) = \sum_{h=1}^{\lceil \frac{l}{r} \rceil} N_c \{1-(h-1)^2 \frac{r^2}{l^2}\}. \tag{4}$$

The total energy E_c required to collect data in the head of a certain cluster is calculated as follows:

$$E_c = x_c(E_{elec}k + \varepsilon_{amp} k r^2) + x_c E_{elec} k. \tag{5}$$

Next, to calculate the energy consumption of a node acting as a cluster head, we derive the energy needed for relaying data. Here, we consider a cluster head in a circular domain of width r_{CH} and whose distance from a sink node is between $d - r_{CH}/2$ and $d + r_{CH}/2$, where r_{CH} is the transmission range of the cluster head. Since most relaying is performed to a cluster head closer to the sink node, we can consider that the probability that relaying will be performed between the cluster heads in this circular domain is very small. Moreover, data from a cluster head outside of the circular domain cannot jump over the domain and be transmitted to the cluster head inside it. Therefore the cluster heads in a circular domain will receive the data generated by all the sensor nodes located outside that domain and will relay it to the sink node or a cluster head closer to the sink node.

Since the areal density of sensor nodes is $N/\pi R^2$ and the areal density of cluster heads is $N_{CH}/\pi R^2$, the amount of data from sensor nodes outside the range of $d + r_{CH}/2 < R$ is

$$\frac{R^2 - (d + r_{CH}/2)^2}{R^2} Nk. \tag{6}$$

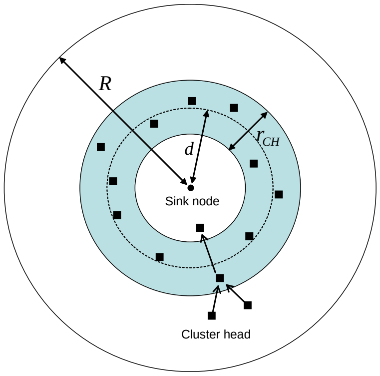


Figure 2. The data generated outside a circular domain is relayed via the cluster heads in that domain.

Moreover, although the number of cluster heads in a circular domain is

$$\frac{(d + r_{CH}/2)^2 - (d - r_{CH}/2)^2}{R^2} \cdot N_{CH}, \tag{7}$$

for simplicity the data relayed from outside that domain shall be equally divided among the cluster heads in the domain. Furthermore, a cluster head combines the data generated by N/N_{CH} nodes in its cluster and transmits that data to the sink node or a cluster head closer to the sink node.

On the other hand, a cluster head din the range of $d + r_{CH}/2 > R$ does not relay data from farther out because there are no sensor nodes farther from the sink node. A cluster head there receives only the data generated in the sensor nodes within its cluster and transmits it to the following cluster head. Moreover, if a cluster is a circular area of radius l, in order that a cluster may stick out of sensing area, the number of sensor nodes in such a cluster is calculated from the following area. The area of a cluster that overlaps in sensing area can be derived as follows:

$$l^2 \cdot \left(\pi - \arccos\left(\frac{x-d}{l}\right)\right) + R^2 \arccos\left(\frac{x}{R}\right) - dy, \tag{8}$$

where

$$x = \frac{R^2 - l^2 + d^2}{2d}, \;\; y = \sqrt{R^2 - x^2},$$

and the number of sensor nodes can be obtained in approximation. Now we can obtain the energy consumption in the case of the usual sensor node and in the case of a cluster head. Since a cluster head is changed periodically in order to avoid failure of its battery, we consider the average consumption energy per sensor node calculated by dividing the total energy consumption by the number of nodes.

The parameter values used in the numerical example are listed in Table 1. Moreover, we compared the results of our analysis with the results of simulation experiments. The results obtained for cluster head transmission ranges of 40, 100, and 200 m are shown in Fig. 3, where the simulation results plotted are the average values for every circular domain width range of 15 m. This result shows that

Table 1. Parameter settings

Number of nodes N	2000
Radius of sensing area R	300 m
Transmission range of sensor nodes r	20 m
Guard zone Δ	none (i.e., Δ =0 m)
E_{elec}	50 nJ/bit
ε_{amp}	100 pJ/bit/m^2

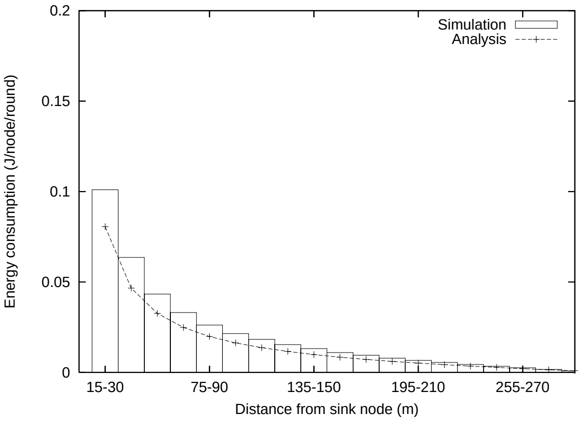
energy consumption of sensor nodes nearer to a sink node increases rapidly. This is because data is accumulated in sensor nodes close to a sink node because data fusion is not performed in this system. Moreover, the energy consumption energy must be large if the communication distance between cluster heads is large (i.e., if the clusters are large). That is, from a viewpoint of energy consumption, smaller clusters are better.

The result of analysis and simulation agree well when the cluster head communication distance is 100 m, but when this distance is large the value obtained by analysis tend to become large and when this distance is small the value obtained by simulation tends to become large. The reasons for this are thought to be the following:

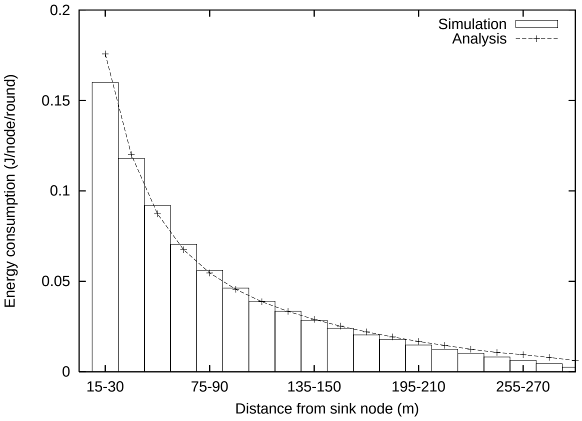
- Although it is assume in analysis that the clusters are circular and contain the same number of sensor nodes, the number of sensor nodes in a cluster actually varies. Since the energy consumption energy of nodes operating as a cluster head is larger rather than that or ordinary sensor nodes, the energy consumption obtained in a simulation becomes large.
- Clusters are assumed to be circular even though actual clusters are not. That is, since distance from the sink node to a sensor node is actually greater than assumed, the energy consumption calculated for nodes near the edge of the cluster artifactually large.

4.2 Total energy consumption of entire network

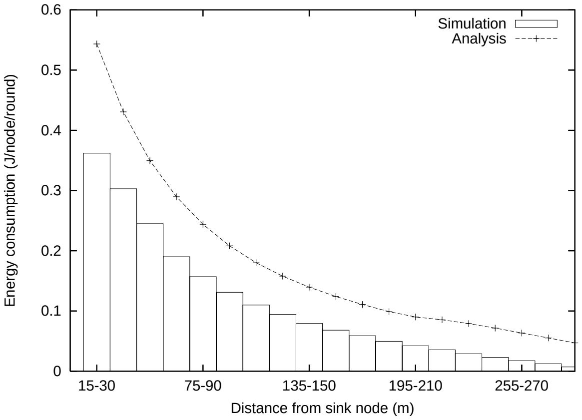
From the relation between energy consumption and the distance from the sink node (obtained for the previous section), we can get the total energy consumption of the whole network. Figure 4 shows how the total energy consumption changes as a function of ε_{amp}, which is a coefficient that shows the distance-dependence in the energy consumption model. Total consumption energy increases rapidly with increasing and becomes remarkable when the distance between cluster heads is large. When ε_{amp} is small, however, the influence of distance will become small. When $\varepsilon_{amp} = 0$, for example, the total energy consumption will be large because the number of times of transmission and reception will increase if a cluster is small.



(a) r_{CH}=40 m



(b) r_{CH}=100 m



(c) r_{CH}=200 m

Figure 3. Relation between average energy consumption and distance from the sink node.

4.3 Performance improvement obtained by not relaying data near the sink node

In a system that does not use data fusion, when multi-hop communication is performed between cluster heads, data will accumulated in the cluster head nearer to the sink node.

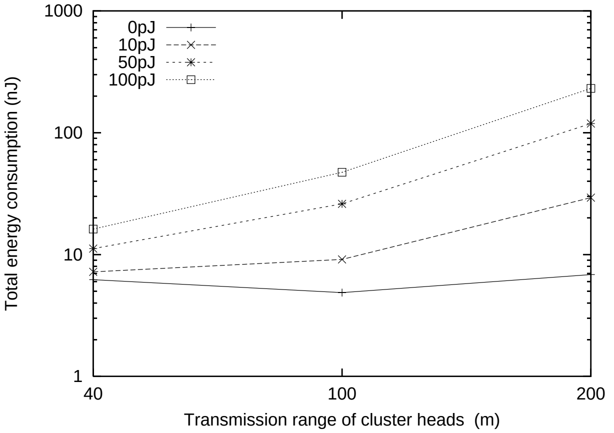


Figure 4. Relation between ε_{amp} and total energy consumption.

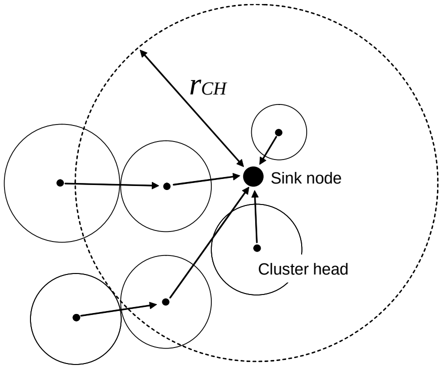


Figure 5. Cluster heads within a radius r_{CH} transmit directly to the sink node.

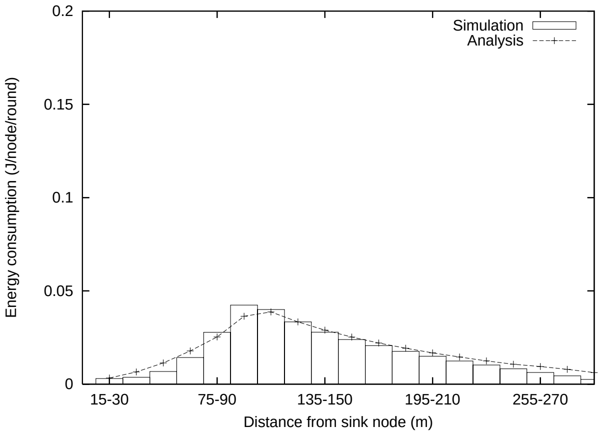


Figure 6. Performance improvement due to not relaying data near the sink node.

Therefore, the amount of data relayed data at the time of becoming a cluster head increases with the proximity of the sensor node to the sink node, and the energy consumption nodes close to the sink node thus becomes very large. So to prolong the life of sensor nodes near the sink node, the relay load should be reduced as much as possible. That is, as shown in Fig. 5, useless relaying can be reduced by transmitting directly to the sink node rather than transmitting to an adjacent cluster between the cluster heads which are within r_{CH} of the sink node. Analysis and a simulation results are shown in Fig. 6, where it can be seen avoiding relaying data within a radius r_{CH} makes the energy consumption in the neighborhood of the sink node very small.

5 Evaluation of the throughput per sensor node

Since the packets in the target sensor network tend to concentrate in the receiving channel of the sink node, that channel acts as a bottleneck. The throughput per sensor node can therefore be obtained as the reciprocal of the schedule length needed to assure that the transmitting channel of the cluster heads transmitting directly to the sink node does not cause interference. Now let n_{1hop} be the number of cluster heads that can transmit to the sink node in one hop:

$$n_{1hop} = \frac{\pi {r_{CH}}^2}{\pi R^2} N_{CH} \tag{9}$$

Here we focus on one certain cluster head. The transmission from this cluster head to a sink node does not receive interference unless there is also transmission from other cluster heads with in the circle of radius $2r_{CH} + \Delta$ centering on this cluster head. If it is assumed that the load on these cluster heads is the same, the schedule length s that does not cause interference will be the number of cluster heads in the circle of radius $2r_{CH} + \Delta$. Therefore, s is obtained as follows:

$$s = \frac{\pi(2r_{CH} + \Delta)^2}{\pi R^2} N_{CH}. \tag{10}$$

Let W denote the transmission capacity of wireless channel, let λ denote the per node throughput, and let all the data generated in the sensing area be transmitted to the sink node through any of n_{1hop} cluster heads. If it is assumed that those loads are distributed equally, we have the following relation:

$$\frac{N\lambda}{n_{1hop}} = \frac{W}{s} \tag{11}$$

Then the per node throughput is given by

$$\lambda = \frac{W}{4n(1 + \Delta(\frac{1}{r_{CH}} + \frac{\Delta}{4{r_{CH}}^2}))}. \tag{12}$$

The results of analysis and simulation are shown in Fig. 7, where the simulation results are the average throughput and standard deviation obtained in 100 simulations based on schedule length until the packet from all sensor nodes reach the sink node. The results obtained when relaying to the cluster heads adjacent to the sink node are

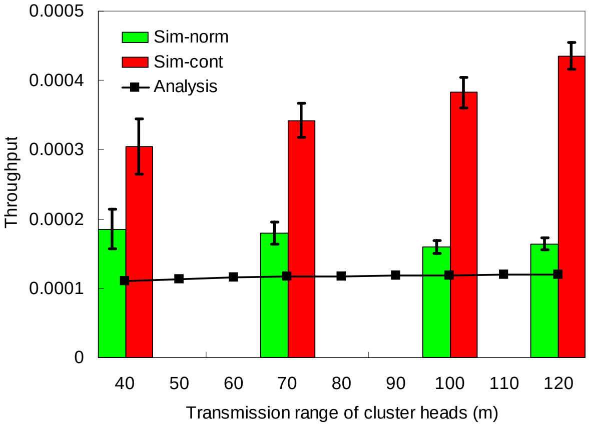


Figure 7. Throughput per node versus cluster head transmission range.

labeled "Sim-norm," and those obtained when transmitting directly to sink node are labeled "Sim-cont." Comparing these results reveals that throughput can be roughly doubled by not relaying data near the sink node. That is, the maximum value 0.0005 (reciprocal with 2000 nodes) for throughput is approached when the cluster radius is close to 300 m. Thus, with regard to cluster size there is a trade-off relation between throughput and energy consumption. Since the interference conditions are conservative for safety, the throughput obtained by analysis is smaller that that obtained in the simulation. That is, scheduling that does not cause interference even if data is transmitted by another cluster head within a radius $2r_{CH} + \Delta$. Therefore, it is possible to collect the data from all sensor nodes by using a short schedule length.

6 Conclusion

In this paper we analytically evaluated the influence of parameters such as the transmission range of cluster heads have on the energy consumption and throughput of a large-scale sensor network that uses multi-hop communication between cluster heads. The throughput was shown to increase with increasing cluster size, but so did the energy consumption energy of the sensor nodes. We also showed that throughput and energy consumption can both be improved very much by controlling the load in the neighborhood of the sink node.

In the target model, all sensor nodes are synchronized and use TDMA communication. In a large-scale actual system, however, such central control is unfeasible. That is, it is necessary to perform dispersive operation based only on the local information near each sensor node. In that case, since interference causes data loss, it will be necessary to evaluate the rate of data collection. Moreover, since some applications require that data be obtained reliably, it is also important to evaluate the increased load or delay caused by retransmission. The influence of the overhead by a control signals required for network operations such as clustering or routing should also be investigated.

Acknowledgment

This research was partially supported by "The 21st Century Center of Excellence Program" and a Grant-in Aid for Scientific Research (C) 17500043 from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

[1] E. J. Duarte-Melo and M. Liu, Data-gathering wireless sensor networks: Organization and capacity, *Computer Networks*, 43(4), 2003, 519–537.

[2] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, *Proc. Hawaii Int. Conf. on System Sciences*, 2000.

[3] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, Data gathering algorithms in sensor networks using energy metrics, *IEEE Trans. Parallel and Distributed Systems*, 13(9), 2002, 924–935.

[4] M. Younis, M. Youssef, and K. Arisha, Energy-aware routing in cluster-based sensor networks, *Proc. Int. Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, 2002.

[5] W. Choi, P. Shah, and S. K. Das, A framework for energy-saving data gathering using two-phase clustering in wireless sensor networks, *Proc. 1st Int. Conf. on Mobile and Ubiquitous Systems: Networking and Services*, 2004.

[6] H. Chan and A. Perrig, ACE: An emergent algorithm for highly uniform cluster formation, *Proc. European Workshop on Wireless Sensor Networks*, 2004.

[7] O. Younis and S. Fahmy, Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach, *Proc. IEEE INFOCOM*, 2004.

[8] B. J. Clupepper, L. Dung, and M. Moh, Design and analysis of hybrid indirect transmission (HIT) for data gathering in wireless micro sensor networks, *ACM Mobile Computing and Communications Review*, 8(1), 2004, 61–80.

[9] P. Gupta and P. R. Kumar, The capacity of wireless networks, *IEEE Trans. on Information Theory*, 46(2), 2000, 388–404.

[10] E. J. Duarte-Melo and M. Liu, Energy efficiency of many-to-one communications in wireless networks, *Proc. IEEE 45th Midwest Symposium on Circuits and Systems*, 2002.