Distance-based Energy Efficient Clustering for Wireless Sensor Networks

SangHak Lee[1,2] JuneJae Yoo[1] TaeChoong Chung[2]

[1] *Ubiquitous Computing Research Center*
Korea Electronics Technology Institute, Korea
{shlee,yoojj}shlee@keti.re.kr

[2] *Department of Computer Engineering*
KyungHee University, Korea
tcchung@khu.ac.kr

Abstract

Sensor Networks have recently emerged as an important computing platform. However, the energy constrained and limited computing resources of the sensor nodes present major challenges in gathering data. In this work, we propose a new distributed clustering and data aggregation algorithm, CODA (Cluster-based self-Organizing Data Aggregation), based on the distance from the sink in ad-hoc wireless sensor networks. While cluster-based data gathering is efficient at energy and bandwidth, it's difficult to cluster efficiently. We use the distance vector from the sink, which affects the energy depletion of the network. Our simulation results show that the proposed algorithm balances the energy dissipation over the whole network thus prolongs the network lifetime.

1. Introduction

The key challenge of gathering and routing data in wireless sensor network is conserving the sensor's energies, so as to maximize their lifetime [1]. The difference between sensor network and standard temperature sensor is that sensor network can able to interconnect nodes intelligently in cluster and to aggregate data collectively [2]. However, distributed self-clustering is a difficult challenge [4]. We will concentrate in this paper on clustering data originating from set of sensors, concurrently operating in the same environment. In this paper, we present a new efficient method for clustering node and aggregating data. The key idea in our algorithm is first to cluster nodes based on the distance from the sink and then to aggregate data using unsupervised learning, the Kohonen Self-Organizing Map (SOM) [5] in each cluster. Our proposed algorithm increases the efficiency (Data received/Total Energy Spent) about 30%.

There are already a lot of works related to cluster for networking [4]. Several heuristics have been proposed to choose clusterheads in ad hoc networks. They include ID-based, connectivity-based and weight-based. Clustering algorithm for ad hoc network is to maintain network topology, while the main goal of sensor network' clustering is to balance energy consumption over the whole network. LEACH [2], ASCENT, SPAN [7], GAF [6], ACE and HEED [4] are tried to preserve and balance the energy dissipation of the network using cluster-based architecture, thus prolong the network lifetime.

Some studies have been done on aggregating data at clusterhead. Madden *et al.* [3] discuss the implementation of traditional database aggregations. In other point of view, aggregating sensor data is classifying multi-input signals into a few classes. The Kohonen SOM (Self-Organizing Map) [5] has a similar characteristic: sensor nodes (referred to as neurons) are recruited topologically for tasks depending on the sensory input. It is commonly classified as a neural network, and more specifically a winner-takes-all competitive algorithm, since the units compete with each other for specific tasks. We adopt the Kohonen SOM to cluster and aggregate sensor data.

2. Clustering and aggregating algorithm

The unbalance of energy depletion is caused by different distance from the sink. We improve the probabilistic fair clustering which is motivated from the LEACH protocol. We divide the whole network into a few group based on the distance from the sink and the strategy of routing. Each group has their own number of cluster number and member node. Then, we apply the scheme to clustering in terms of the routing strategy.

$$k_j = k * \frac{D_j}{\sum_{j=1}^{n} D_j} \quad (1 \le j \le n) \tag{1}$$

where D_j is the distance from the sink and smaller index means closer to the sink. K is the number of clu-

Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks (LCN'04)
0742-1303/04 $ 20.00 IEEE

ster in each group.

In case of multi-hop with clustering as follows:

$$k_j = k * \frac{D_{n-j+1}}{\sum_{j=1}^{n} D_j} \quad (1 \le j \le n) \qquad (2)$$

Eventually, we abstract the probabilistic cluster decision function for each group as follow:

$$P_i(t) = \begin{cases} \frac{k_j}{N_j - k_j * \left(r \bmod \frac{N_j}{k_j} \right)} : C_i(t) = 1 \\ 0 : C_i(t) = 0 \end{cases} \quad (v_i \in G_j) \qquad (3)$$

We propose an algorithm which trains the sensor nodes using the Kohonen SOM to aggregate data. The Kohonen SOM has three phases: Initialization, Winner Finding and Weights Updating. The original SOM has an array of neurons that compute simple output functions of incoming inputs of arbitrary dimensionality, a mechanism for selecting the neuron with the largest output, and an adaptive mechanism that updates the weights of the selected neuron and its neighbors. There are some different requirements that we have to modify from the original Kohonen SOM. Since sensor network may consist of hundreds or thousands of nodes, it is difficult to implement in centralized method. Our idea is to classify the sensed data as well as to re-cluster the self-organizing map using the Kohonen SOM. We choose a cluster head which manage the communication schedule, merge and partition the clusters. There are geographical clusters right after deploying sensor network and then clusters are merged and partitioned based on data after operating the Kohonen SOM. Data-centric cluster scheme can achieve the accuracy of aggregated data and thus make user to become aware the situation of sensor field.

3. Simulation results and conclusions

We evaluate the performance of the CODA algorithm via simulations. In our experiments we used mainly two metrics: *energy consumption and quality of data*. We compare CODA to a probabilistic even clustering protocol that is for balancing energy consumed. We use the network in which 100 nodes are located randomly in 100×100 dimensions. What we simulate to verify the idea related with energy consumption is the number of node alive and the amount of data received at the sink over time. Our clustering algorithm, CODA, increases the network lifetime about 30% over LEACH.

We inspect whether distance-based clustering scheme affect the quality of data. Thus we take a look at the quality of data in a cluster. There are five units in a cluster where they aggregate data using different methods: AVERAGE and SOM. They sense different value from the same event. We measured the difference between original values and aggregated using light sensor. Since the aggregation methods used in database system have no knowledge in deciding if the heterogeneous data should be included in one data set, it's hard to partition the dataset. Simulation results verify that aggregating data based on machine learning is effective.

In this paper, we present a self-organizing clustering method based on the distance from the sink and an aggregating data using competitive machine learning. We differentiate the number of cluster in terms of the distance from the sink. The farther the distance from the sink, the more cluster is formed in case of single hop with clustering. It shows better performance than applying the same probability to the whole network in terms of the network lifetime and the dissipated energy.

In addition, we use the unsupervised self-organizing neural network, Kohonen SOM, to map sensor data to meaningful information. Our proposed algorithm, the CODA, works well in increasing the quality of data and decreasing the energy consumption to transmit data in simulation.

4. References

[1] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks", *MobiCOM '99*, Seattle, WA., Aug. 1999, pp. 263-270.

[2] 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, no. 4, Oct. 2002, pp.660-670.

[3] S. Madden, R. Szewczyk, M. J. Franklin, and D. Culler, "Supporting Aggregate Queries over Ad-Hoc Wireless Sensor Networks.", *Proc. of 4th IEEE Workshop on Mobile Computing and Systems Applications*, June 20-21, 2002, pp.49-58.

[4] O. Younis and S. Fahmy, "Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach", *IEEE INFOCOM 2004*, Mar. 2004.

[5] T. Kohonen, "The self-organizing map", *Proceedings of the IEEE*, Sept. 1990. pp.1464-1480.

[6] Y. Xu, J. Heidemann, and D. Estrin, "Geography-Informed Energy Conservation for Ad Hoc Routing," MOBICOM 2000, Rome, Italy, July 2001, pp. 70-84.

[7] B. Chen, K. Jamieson, H. Balakrishnan and R. Morris, "Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks", *ACM Wireless Networks Journal*, Vol.8, No. 5, Sep. 2002, pp.481-494.