Inner-Cluster Data Aggregation Scheme for Wireless Sensor Networks

Jianghong Guo, Yudong Luo, Deli Chen,

School of computer, JiaYing University

Meizhou, China

Email: g_jh@jyu.edu.cn, mzlyd@jyu.edu.cn, dlchen@jyu.edu.cn

Abstract— **For reducing the communication overhead of data aggregation, authors proposed an inner-cluster data aggregation scheme for wireless sensor networks. Agent nodes are selected with assist of location information and only the nodes have different readings with agent nodes transmit their data to cluster head, thus reduce the inner-cluster transmissions. Analysis and experiments show that our scheme reduces inner-cluster transmissions effectively and lowers the communication overhead.**

Keywords- sensor network; data aggregation;agent nodes

I. INTRODUCTION

Wireless Sensor Network consists of a large number of resources limited sensors and one of the important tasks of WSN is collecting and transmitting the surrounding parameters to base station. Usually, sensors are deployed in hostile environment and it is difficult to replacing batteries, thus energy consumption becomes one of the most important considerations of protocol designing.

In WSN, sensors communicated with each other using wireless signal and one transmitted packet will received by all of the sender's neighbors, thus the communication overhead is the major consumption of sensor's energy. Data aggregation is one of the most important methods for saving energy and many schemes had been proposed for decreasing the redundant transmissions.

Al-Karaki et al. proposed exact and approximate aggrega-tion algorithms[1]; Aonishi et al. studied the impact of agg- regation efficiency on GIT routing [2]; Villas et al. proposed a scalable and dynamic data aggregation aware routing protocol for wireless sensor networks [3]; Heinzelman et al. proposed LEACH [4] for the network clustering and cluster head election; Younis et al. proposed HEED[5] in which the residual energy was taken into account cluster head election; NECCHI et al. applied Gossip algorithm to data aggregation [6] and this method was improved in [7]; FAN et al. proposed an efficient and robust sensor data aggregation using linear counting sketches[8].

Usually, the cluster-based data aggregation consists of two parts: aggregation of inner-cluster readings and transmitting fusion data to base station. Generally, there are redundant data in WSN because neighbor sensors may have same or similar readings, but in most of existing schemes，all inner-cluster sensors will transmit their data to the aggregator for further processing. Authors proposed a inner-cluster data aggregation scheme (ICDA) and reduce inner-cluster trans- missions with assistant of sensors' location information.

II. OUR WORK

In our scheme, the cluster will be divided into 4 regions and some nodes selected as reporters for each region. Only the nodes located in the same region and have different readings with reporters will transmit data to the aggregator, thus reduce inner-cluster transmissions effectively.

A. Network model and assumptions

As shown in Fig.1, the network was clustered and we have some assumptions as follow:

1) The network was clustered and each inner-cluster node was 1-hop away from the cluster head, also, the cluster head will responsible for the inner-cluster data aggregation;

2) Due to random scattering, sensors can achieve their locations by some algorithms [9][10] upon deploying and all sensors have same communication radius R (R=40m);

3) Assume that inner-cluster parameters are follow normal distribution with mean μ and standard deviation σ due to the R is relatively small, also, sensor readings are rounded to integer because the measuring errors are unavoidable.
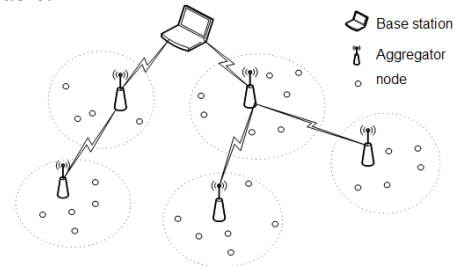

Figure 1. Network model

B. Cluster division

Assume that there are M sensors deployed in the network, each sensor, N_i, $i \in [1,M]$, can achieves its location (X_{Ni}, Y_{Ni}) and the position of cluster head Ch. Then, the cluster will be divided into 4 regions as follow:

As shown in Fig.2 (a), the cluster head Ch locates in the center and the communication area of Ch was divided into 4 regions which are ranged from $(0, \pi/4)$, $(\pi/4, \pi/2)$, $(\pi/2, 3\pi/2]$ and $(3\pi/2, 2\pi]$ respectively. Assume that region S1 is sector OAB, SA is the circle with point A as center and R as radius, SB is the circle with point B as center and R as radius. The report region SR of S1, the dark shadow area shown in Fig.2

(a), satisfies SR=S1∩SA∩SB. Also, we denote "o" as the nodes outside SR and "*" as the nodes within SR as shown in Fig.2(b).

Theorem 1：If SR is the report area of region S, for any point V located in SR and U located in S, the distance from V to U less than or equal to R, R is communication radius.

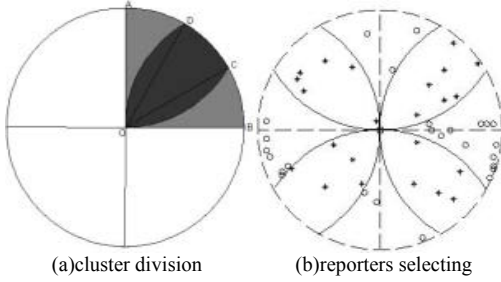Theorem 1 can be deduced with plane geometry easily, so we omit the proof.

(a)cluster division   (b)reporters selecting

Figure 2.   inner-cluster partition and reporters selecting

C.   Determining the reporter node

Assume that the location of cluster head Ch is (X_{Ch}, Y_{Ch}) and the location of 4 split points are (X_A, Y_A), (X_B, Y_B), (X_C, Y_C) and (X_D, Y_D) respectively, Ch broadcasts the locations to all inner-cluster nodes. For each inner-cluster node Ni, it calculates the located region as follows:

$$\theta = \arctan(\frac{X_{Ni} - X_{Ch}}{R}) \tag{1}$$

Assume that $\theta \in [0, \pi/2]$, then：

$$\begin{cases} if \ X_{Ni} > X_{Cn} \ , \ Ni \in S1 \\ if \ X_{Ni} < X_{Ch}, \ Ni \in S3 \end{cases} \tag{2}$$

In the similar way, Ni can determine if it locates in $S2$ or $S4$.

The node is a candidate reporter if it located in the report area. Assume that node Ni located in $S1$ and the distances from Ni to Ch, A and B are d_{Ch}, d_A and d_B respectively, then Ni is a candidate reporter when equation (3) holds. Ni transmits $<Ni, S1,1>$ or $<Ni, S1,0>$ to cluster head where 1 means Ni within SR and 0 means Ni outside SR. Upon receiving such messages, the cluster head can learn the number of nodes located in each region and if the nodes are candidate reporters.

$$d_{Ch} \leq R \ and \ d_A \leq R \ and \ d_B \leq R \tag{3}$$

TABLE I.   REPORTER INFORMATION

ID	Area	Num
N_1	$S1$	Num_1
N_2	$S2$	Num_2
N_3	$S3$	Num_3
N_4	$S4$	Num_4

For each region, cluster head selects randomly one of the candidate reporters as reporter node and preserves reporter information as shown in Table I. Where the count means the number of nodes located in the region, for example, Num_1 means the number of nodes located in $S1$. Also, none nodes located in the report area of $S1$ will lead to none reporter in $S1$. Then, cluster head broadcasts the reporter list $<N1, S1, N2, S2, N3, S3, N4, S4>$ to inner-cluster nodes.

D.   Data aggregation

Upon receiving the data request of BS, sensors detect sur-rounding parameters and reporters transmit their readings to cluster head Ch, Ch aggregates reporter readings and save the results in aggregation table (TA) as shown in Table II. Where Data is the reporter's reading and Num is the number of nodes located in the corresponding region.

For any non-reporter node Ni, Ni compares its reading with that of the corresponding reporter and keeps silent if its reading is the same with reporter's reading, otherwise Ni will transmits $<ID, region, data>$ to Ch.

For example, Ni locates in $S1$ and has different reading with that of N_1, Ni will transmits $<Ni, S1, D_{Ni}>$ to aggregator. For convenience, we call the jth line in TA $record(j)$ and denote $record(j)->*$ as the corresponding item of this record. Upon receiving the message $<Ni, S1, D_{Ni}>$, aggregator works as follow:

For all records with region $S1$, aggregator compares their data with D_{Ni}. If there exists $record(j)$ whose data equals D_{Ni}, then $record(j)->num$ plus 1, also, $record(1)->num$ minus 1 because Ni locates in $S1$ and has different reading with the reporter N_1; otherwise $<Ni, S1, D_{Ni}>$ will be inserted into TA as a new record. Other messages, such as $<Nj, S2, D_{Nj}>$, will be processed in the similar way.

Therefore, the aggregator can learn different data and corresponding count of all regions. Based on TA, final aggregation results can be achieved.

TABLE II.   TABLE OF AGGREGATION (TA)

ID	Area	Data	Num
N_1	$S1$	D_N_1	Num_1
N_2	$S2$	D_N_2	Num_2
N_3	$S3$	D_N_3	Num_3
N_4	$S4$	D_N_4	Num_4

Because main concern of this paper is to reduce the inner-cluster transmissions and many existing methods can be used for transmitting fusion data to the base station, thus we don't detail this issue.

III.   ANALYSIS AND SIMULATION
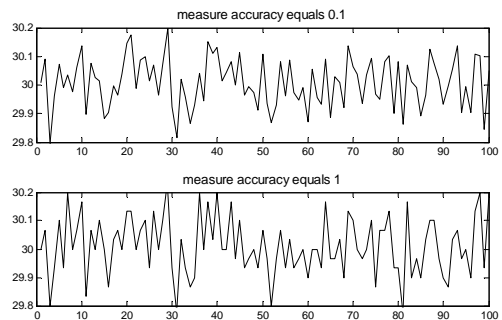
A.   Analysis

Figure 3.   inner-cluster means with different measure accuracy

Authors use random numbers of normal distribution to simulate the readings of temperature sensors. Assume that sensor readings follow (30, 0.5) normal distribution where 30 is mean and 0.5 is standard deviation. We generate 30 simulation data and calculate their means with different measure accuracy (0.1 and 1). Upon repeating the experiment 100 times, we have the results as shown in Fig.3.

From Fig.3 we know that the inner-cluster means with different measure accuracy (0.1 and 1) are both ranged from [μ-2, μ+2], at the same time, the measurement error of the sensor is unavoidable, thus, rounds the sensor readings to integers has effect on calculating inner-cluster means limitedly.

Assume that the set of n sensor readings is {$Data_i$}，$i \in$ [$1,n$], the probability for sensor reading equals $Data_i$ is p_i. Thus, the probability P for any node which has the same reading with the corresponding reporter satisfies equation (4).

<div style="text-align:center">错误!未找到引用源。</div>

$$P = \sum_{i=1}^{n} p_i^2$$

<div style="text-align:center">(4)</div>

If there are N inner-cluster nodes, the number of nodes which transmit their data to aggregator is approximately equals:

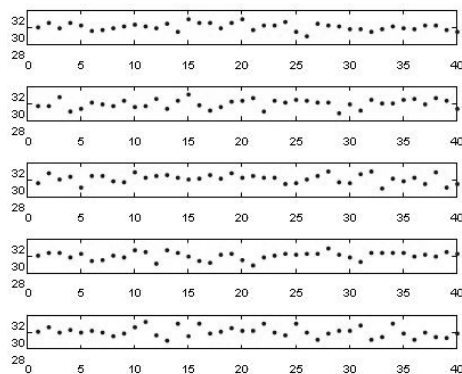$$N_{trans} = 4 + (N-4) \times (1-P) \qquad (5)$$

B. Simulation

Figure 4.   The data graph of 5 clusters in one experiment

We use CDA (Cluster-based Data Aggregation) denotes most of existing cluster-based schemes (such as [6], [7], [8]) in which all inner-cluster sensor readings will be transmitted to the aggregator.

The adopted network model as shown in Fig.1 and assume that number of nodes in each cluster is 40, we generate random numbers of normal distribution for all of nodes and Fig.4 shows the data graph of 5 clusters in one experiment.

The comparison of inner-cluster transmissions with ICDA and CDA was shown in Fig.5, where ICDA denotes the number of inner-cluster transmissions in one experiment with data graph shown in Fig.4 and ICDA* denotes the average number of inner-cluster transmissions on 50 trails.

From Fig.5 we know that the inner-cluster transmissions in ICDA are about 32.5%，57.5%，60%，70% and 62.5%

of that in CDA respectively, and the inner-cluster transmissions in ICDA* are about 52.5%，50%，52.5%，55% and 52.5% of that in CDA respectively.

It is obvious that the inner-cluster transmissions in our scheme is lower than that of CDA, the reason is that only the nodes which have different readings with their reporters transmit data to aggregator in our scheme, but all inner-cluster nodes will send their readings in CDA.

Upon rounding the sensor readings into integers, the readings set {$Data_i$} is {28, 29, 30, 31, 32} and the corresponding probability for sensor reading equals $Data_i$ is 0.13%, 15.51%, 68.72%, 15.51% and 0.13% respectively. According to equation (5), the average number of inner-cluster transmissions is 21 theoretically. Average number of inner-cluster transmissions of 5 clusters in ICDA* are 21, 20, 21, 22 and 21 respectively, It shows that the results of our method are very close to the theoretic results and the proposed scheme is feasible.
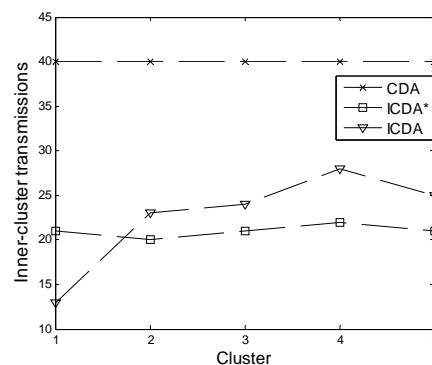

Figure 5.   inner-cluster transmissions

IV.   CONCLUSION

In this paper, we proposed a cluster trisecting based data aggregation scheme for cluster-based data aggregation. Cluster was divided into 4 regions and all nodes determined the located region with assistant of location information. Only the nodes, which located in the same region and have different reading with corresponding reporter, will transmit their data to aggregator. Simulation shows that our scheme reduces the inner-cluster transmissions effectively, thus lowers the communication overhead and prolongs the network lifetime.

REFERENCES

[1] Intanagonwiwat, C Govindan, R Estrin, D Heidemann. "Directed diffusion for wireless sensor networking", In IEEE transactions on networking, 2003, 11(1): 2–16.

[2] T. Aonishi, T. Matsuda, et al. "Impact of aggregation efficiency on GIT-routing for wireless sensor networks", in Proc of the 2006 International Conference Workshops on Parallel Processing, Ohio, USA, IEEE Society Press, 2006:151–158.

[3] L. A. Villas, D. L. Guidoni, R. B. Ara´ujo, A. Boukerche. "A scalable and dynamic data aggregation aware routing protocol for wireless sensor networks", in 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems, USA, IEEE Society Press, 2010: 110–117.

[4] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", IEEE Trans Wireless Communication. Oct, 2002: 660–70.

[5] O. Younis and S. Fahmy, "HEED: a Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor networks", IEEE Trans. Mobile Computing, 2004, 3(4): 366–79.

[6] NECCHI L, BONIVENTO A, LAVAGNO L, et al. "Eerina: an energy efficient and reliable in-network aggregation for clustered wireless sensor networks", in Proc of Wireless Communications and Networking Conference, 2007: 3364-3369.

[7] MOSTAFIZUR M，MOZUMDAR R, et al. "An efficient data aggregation algorithm for cluster- based sensor network", Journal of Networks, 2009, 4(7) : 598-606.

[8] FAN Y C et al,"Efficient and robust sensor data aggregation using linear counting sketches", IEEE Trans on Parallel and Distributed Systems, 2010, 21(11) :1675-1691.

[9] S.Capkun, J.P. Hubaux, "Secure positioning of wireless devices with application to sensor networks", in *Proc. IEEE INFOCOM*, Miami, FL, Mar. 2005, pp. 1917– 1928.

[10] A. Savvides, C. Han and M. Srivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors", in *Proc. ACM MobiCom*, Rome, Italy, Jul. 2001, pp. 166–179.