Proceedings of the 2012 2nd International Conference on Computer and Information Application (ICCIA 2012)

Cluster Trisecting based Data Aggregation Scheme for Wireless Sensor Networks

Jianghong Guo
School of computer, JiaYing University
Meizhou, China
e-mail: g_jh@jyu.edu.cn

Deli Chen
School of computer, JiaYing University
Meizhou, China
e-mail: chendl@jyu.edu.cn

***Abstract*—Data aggregation is an important method to reduce energy consumption in wireless sensor networks (WSN). Authors proposed a cluster trisecting based data aggregation scheme for wireless sensor networks in which the cluster was trisected and some reporters were assigned to each region. The nodes have same reading and located in same region with reporter will keep silent in data aggregating, thus reducing the inner-cluster transmissions. Analysis and simulation show that the transmissions of inner-cluster aggregation in our scheme lower than that of related schemes and the decrease of transmissions is obvious when redundancy of sensor readings is high.**

Keywords- sensor network; data aggregation, cluster trisecting

I. INTRODUCTION

Wireless Sensor Network consists of a large number of resources limited sensors which are deployed by stochastic scattering or deterministic deploying. One of the important missions of WSN is collecting the surrounding parameters and transmitting these data 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 the data transmitted by a sensor will received by all neighbors, thus the communication overhead is the major consumption of sensor's energy. Data aggregation is one of the most important methods to reduce the communication overhead and many schemes had been proposed for decreasing the redundant transmissions.

Al-Karaki et al. proposed exact and approximate aggregation algorithms[1]; Aonishi et al. studied the impact of aggregation 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: inner-cluster aggregation and fusion data transmission. We know that there are redundant data in WSN because neighbor sensors may have same or similar readings, but all inner-cluster sensors will transmit their data in most of existing schemes. Authors proposed a cluster trisecting based data aggregation scheme(CTDA) and reduce the inner-cluster transmissions with assistant of location information.

II. OUR WORK

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

A. Network model and assumptions

In this paper, we have some assumptions about the network. 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. The network model was shown in Fig.1; 2) Due to random scattering , sensors achieve their locations by some location algorithms [9][10] upon deploying; 3) there are multi-type sensors located in the deployed area to detect various parameters and all sensors have same communication radius *R* (*R*=40m); 4) we assume that inner-cluster parameters are follow normal distribution with mean μ and standard deviation σ due to the communication range of sensor is relatively small, sensor readings are rounded to integer because the measuring errors are unavoidable.

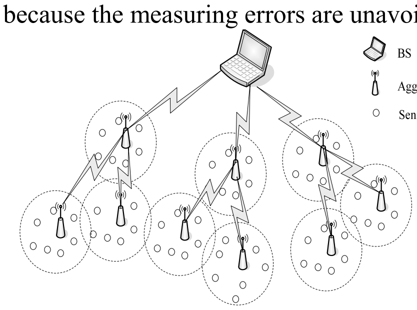


Figure 1. Network model

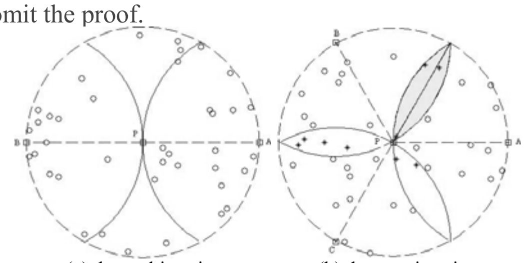
B. Cluster trisecting

Assume that there are *x* sensors deployed in the network, each sensor, *Ni*, $i\in[1,x]$, can achieves its location (X_{Ni},Y_{Ni}) and the position of cluster head *P*. Then, the cluster head *P* will trisect the cluster as follow:

As shown in Fig.2 (b), the cluster head *P* calculates the location of split point *A, B* and *C* and divides the communication area into trisections *S1*, *S2* and *S3* which are ranged from $(0,2\pi/3]$, $(2\pi/3,4\pi/3]$ and $(4\pi/3, 2\pi]$ respectively. Assume that *S1* is sector *PAB*, *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 shaded area shown in Fig.2 (b), satisfies $SR=S1\cap SA\cap SB$. Also, we denote “o” as the nodes outside *SR* and “*” as the nodes within *SR* in Fig.2.

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 bisecting (b)cluster trisecting

Figure 2. inner-cluster division

It is obvious that the number of reporters increased with the increase of the number of regions, so we adopt as few regions as possible. If the cluster is one region, only the communication range of cluster head can cover the cluster and it makes the aggregation inefficient when there are multi-type sensors. We adopt cluster trisecting in our scheme because the case of cluster bisecting is similar with one region division as shown in Fig.2 (a).

C. Determining of candidate reporter

Assume that the location of cluster head P is (X_P, Y_P) and the location of split points are (X_A, Y_A), (X_B, Y_B) and (X_C, Y_C) respectively, P broadcasts the locations to all inner-cluster nodes. For each inner-cluster node *Ni*, it calculates the located region as follows:

$$\theta = \arccos(\frac{X_{Ni} - X_P}{R}) \tag{1}$$

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

$$\begin{cases} if\ Y_{Ni} > Y_P\ and\ \theta < 2\pi/3,\ \mathrm{Ni} \in S1 \\ if\ Y_{Ni} > Y_P\ and\ \theta \ge 2\pi/3,\ \mathrm{Ni} \in S2 \\ if\ Y_{Ni} < Y_P\ and\ \theta < 2\pi/3,\ \mathrm{Ni} \in S3 \\ if\ Y_{Ni} < Y_P\ and\ \theta \ge 2\pi/3,\ \mathrm{Ni} \in S2 \end{cases} \tag{2}$$

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 *P, A* and *B* are d_P, d_A and d_B respectively, then *N*i is a candidate reporter when equation (3) holds. *Ni* transmits <*Ni, S1, type,* 1> or <*Ni, S1, type,0*> to cluster head *P* where type is the type of sensors and 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 which type these sensors are.

$$d_P \le R\ and\ d_A \le R\ and\ d_B \le R \tag{3}$$

D. selection of region reporter

Assume that there are two types of sensors, type *A* and type *B*, deployed in the network. For each type of sensors, cluster head selects randomly one of the corresponding candidate nodes as reporter for each region and preserves reporter information as shown in Table I. Where the count means the number of certain type of nodes located in the region, for example, $C_{A,S1}$ means the number of nodes of type *A* located in *S1*. Also, all nodes of type *B* outside of report area of *S1* will lead to no reporter of type *B* located in S1. Then, cluster head broadcasts the reporter list <*N1, A, S1*>, <*N2, A, S2*>, <*N3, B, S2*>… to inner-cluster nodes.

TABLE I. REPORTER INFORMATION

ID	type	region	count
N1	*A*	*S1*	$C_{A,S1}$
N2	*A*	*S2*	$C_{A,S2}$
N3	*B*	*S2*	$C_{B,S1}$
N3	*A*	*S3*	$C_{A,S3}$
N3	*B*	*S3*	$C_{B,S3}$

E. Inner-cluster data aggregation

Upon receiving the data request of *BS*, sensors detect surrounding parameters and reporters transmit their readings to cluster head *P*, *P* aggregates reporter readings and save the results in aggregation table (*TA*) as shown in Table II. 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, type, region, data*> to *P*. For example, *Ni* is *A*-type sensor located in *S1* and has different reading with that of *N1*, *Ni* will transmits <*Ni, A, 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, A, S1,* D_{Ni}>, aggregator works as follow:

For all records with type A and region *S1*, aggregator compares their data with D_{Ni}. If there exists *record(j)* whose data equals D_{Ni}, then *record(j)->count* plus 1; otherwise <*Ni, A, S1, 1,* D_{Ni}> will be inserted in *TA* as a new record. Other messages, such as <*Nj, B, 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 III is the final result of inner-cluster aggregation and <*A,* C_{A1}, D_{N1}> means there are C_{A1} nodes which are *A*-type sensors and have reading D_{N1}.

TABLE II. TABLE OF AGGREGATION (TA)

ID	type	region	count	data
N1	*A*	*S1*	$C_{A,S1}$	D_{N1}
N2	*A*	*S2*	$C_{A,S2}$	D_{N2}
N3	*B*	*S2*	$C_{B,S1}$	D_{N3}
N3	*B*	*S3*	$C_{A,S3}$	D_{N4}
N3	*C*	*S3*	$C_{B,S3}$	D_{N5}

TABLE III. THE FINAL AGGREGATION RESULTS

type	count	Data
A	C_{A1}	D_{N1}
A	C_{A2}	D_{N2}
...	...	...
B	C_{B1}	D_{N4}
B	C_{B2}	D_{N3}

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

III. ANAYSIS AND SIMULATION

In our scheme, the network is clustered as shown in Fig.1 and related assumptions had been described in section 2.1. We use inner-cluster transmissions to evaluate the communication overhead and use ns2 to simulate the network with wireless extension. Also, *802.15.4* is adopted as communication protocol which allows a variable payload of up to 102 bytes. We assume that the node density (the average number of nodes located in the cluster) is $2k$ (k sensors of type *A* and k sensors of type *B*), means of inner-cluster data distribution are 20 and 40 for two kinds of sensors respecttively and standard deviation σ various from 0.6 to 1. We use CDA (Cluster-based Data Aggregation) denotes most of existing cluster-based schemes in which all inner-cluster sensor readings will be transmitted to the aggregator. The data were rounded to integer and ranged from μ-4 to $\mu+4$ because the probability $P(x \le \mu-4)$ and $P(\mu+4 \le x)$ are very small in normal distribution when $\sigma \le 1$. For any data Di, $Di=\mu-5+i$ and $i \in [1,9]$, the probability P[sensor reading is Di]$=pi$. Thus, the probability P for any node which has the same reading with the corresponding reporter satisfies equation (4).

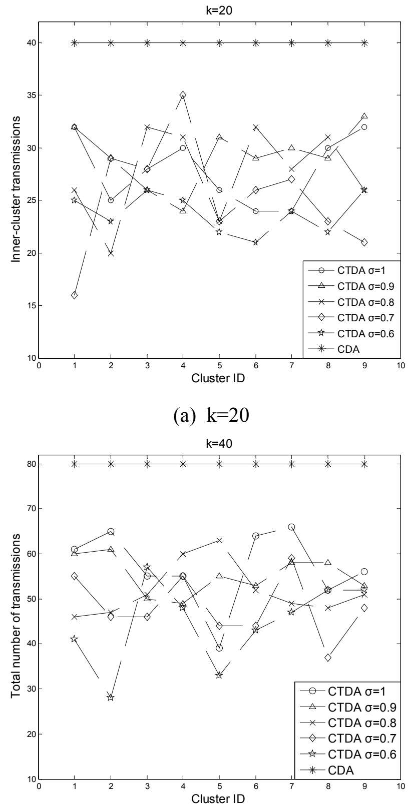
$$P = \sum_{i=1}^{9} (pi)^2 \tag{4}$$

For example, if there are L *A*-type sensors located in $S1$, it is theoretically $(L-1)*P$ sensors have the same reading with the reporter $N1$.

In our simulation, the comparison of inner-cluster transmissions with different k were shown in Fig.3 and total number of inner-cluster transmissions with different σ was shown in Fig.4. We denote Nt-CTDA(sim) as the number of transmissions in simulation and Nt-CTDA(tho) as theoretical value.

From Fig.3 and Fig.4 we know that the decrease of σ and the increase of node density results more redundant data, therefore reduce the inner-cluster transmissions effectively. The difference between Nt-CTDA(sim) and Nt-CTDA(tho) decreased with the increase of node density because there may be no reporter located in the report area of the region and data redundancy is low with low-node density, thus all nodes located in the region will transmit data to aggregator.

Compared with CDA, the inner-cluster transmissions are reduced and the decrease is obvious when redundancy of sensor readings is high. For example, when σ=0.6 and k=40, there are approximately 41.13 percent of the nodes will keep silent in inner-cluster data aggregation.



(a) k=20

(b) k=40

Figure 3. inner-cluster transmissions

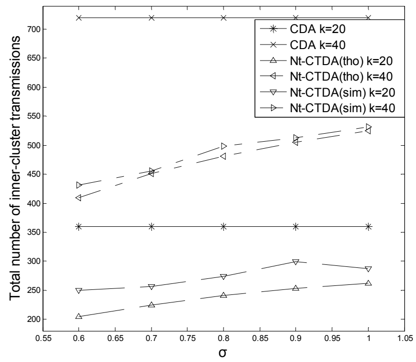


Figure 4. total number of transmissions

IV. CONCLUSION

In this paper, we proposed a cluster trisecting based data aggregation scheme for cluster-based data aggregation. Cluster was trisected 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.