Probabilistic Transmitting-based Data Aggregation Scheme for Wireless Sensor Networks

Jianghong Guo [1, a], Yudong Luo [1,b]

[1] School of computer, JiaYing University, Meizhou, China

[a] g_jh@jyu.edu.cn , [b] mzlyd@jyu.edu.cn

Abstract. For reducing the communication overhead of data aggregation, authors proposed a probabilistic transmitting-based data aggregation scheme for wireless sensor network. Due to limited number of nodes in the cluster and aggregation error is unavoidable; nodes adopt the probabilistic transmitting to lower the communication overhead with tolerable error. In order to prevent the influence of big noise, Dixon criterion was adopted to eliminate the gross error in the small sample. Experiment shows the proposed scheme reduces the inner-cluster transmissions and lowers the communication overhead effectively with tolerable error. The aggregation error of probabilistic transmitting and all nodes transmitting are at the same level and both are acceptable for wireless sensor networks.

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 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].

In wireless sensor networks, certain range of errors is acceptable and it can be used to reduce the communication overhead. But most of existing cluster-based aggregation schemes mainly focuses on how to reduce the energy consumption of fusion data transmission and take little care of reducing the inner-cluster communication overhead. For convenience, we denote the schemes in which all sensors will transmit their readings as traditional cluster-based data aggregation (TCDA).

In this paper, authors proposed a probabilistic transmitting-based data aggregation scheme for WSN and reduced the inner-cluster transmissions under the precondition of tolerable errors.

Our work

A. Network model and assumptions

We have some assumptions about the network. 1) Upon deployment, the network was clustered and the cluster head will responsible for the inner-cluster data aggregation. The network model was shown in Fig.1; 2) 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; 3) All

sensors have same communication radius R (R=40m) . Each sensor transmits its reading to aggregator with the probability P and the reading is disturbed by the big noise with the probability Pn.
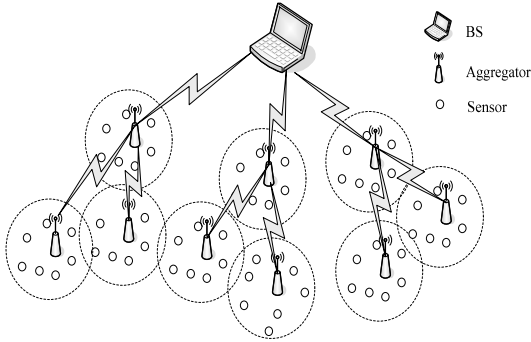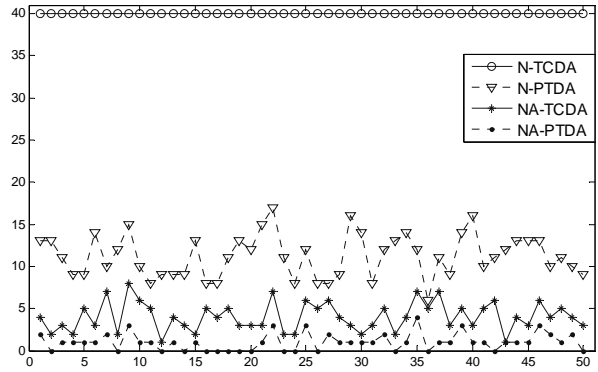

Fig.1 Network model

Fig.2 The number of inner-cluster transmissions and included abnormal readings

B. The reliability of probabilistic transmitting

Neighbor sensors may detect the same or similar data. Sensor's measurement error is unavoidable and the tolerable error may range from [-1%, 1%], thus, we take the tolerable error as the acceptable threshold. For example, assume μ=30，the acceptable threshold ranges from -0.3 to 0.3. According to the 3σ criterion, the probability of the sensor reading belongs to [29.7, 30.3] is 99.73%.

Assume that there are N nodes in a cluster, Di, the reading of node Si, belongs to [29.7, 30.3]. Take average value of K sensor readings for example, the aggregation result RS satisfies

$$RS = (\sum_{i=1}^{K} D_i) / K \in [29.7, 30.3]$$ (1)

It is obvious that K has little impact on RS when sensor readings are reliable. The size of sample set is mainly used to detect the gross error and improve the aggregation accuracy. When the required aggregation accuracy is not very high, we can adopt probabilistic transmitting to reduce the energy consumption with the precondition of required data reliability. The main weakness of probabilistic transmitting is that it will suffer more from big noise than all node transmitting. This means that the probabilistic transmitting requires the gross error of the small sample set should be detected effectively.

C. Gross error detection

In common gross error detection schemes, such as 3σ criterion and Grubbs criterion, it is required the standard deviation must be calculated at first. The calculation needs multiple physical quantities and causes inconvenience. Thus, we adopt the Dixon criterion to detect the gross error. Assume $\{Xi\}$ is the order statistics of normal distribution and $i \in [1,n]$, constructs the statistics as equation (2). The gross error of the sample set can be detected according to equation (3). Where α is significant parameter, n is the size of sample set and $D(\alpha, n)$ is the corresponding critical value.

$$\begin{cases} r_{10} = \dfrac{x_n - x_{n-1}}{x_n - x_1}, r'_{10} = \dfrac{x_2 - x_1}{x_n - x_1} & n = 3 \sim 7 \\[2mm] r_{11} = \dfrac{x_n - x_{n-1}}{x_n - x_2}, r'_{11} = \dfrac{x_2 - x_1}{x_{n-1} - x_1} & n = 8 \sim 10 \\[2mm] r_{12} = \dfrac{x_n - x_{n-2}}{x_n - x_2}, r'_{12} = \dfrac{x_3 - x_1}{x_{n-1} - x_1} & n = 11 \sim 13 \\[2mm] r_{13} = \dfrac{x_n - x_{n-2}}{x_n - x_3}, r'_{13} = \dfrac{x_3 - x_1}{x_{n-2} - x_1} & n = 14 \sim 30 \end{cases}$$ (2)

$$\begin{cases} r_{ij} > r'_{ij} \ and \ r_{ij} > D(\alpha, n)，then \ x_n \ is \ abnormal \ data \\ r_{ij} < r'_{ij} \ and \ r_{ij} > D(\alpha, n)，then \ x_1 \ is \ abnormal \ data \end{cases}$$ (3)

D. Probabilistic transmitting-based data aggregation

The proposed scheme (Probabilistic transmitting-based data aggregation scheme, PTDA) mainly includes four parts as follow:

1) Usually, sensors are deployed by stochastic scattering and it is difficult to pre-determine the number of nodes in a cluster, thus, we must preload multiple critical value of gross detection for different size of sample set before deployment. Assume that α is significant parameter, n_i is the size of sample set, $D(\alpha, n_i)$ is the corresponding critical value and $n_i < n_{i+1}$, parameters $<n_1, D(\alpha, n_1)>$, $<n_2, D(\alpha, n_2)>$, .., $<n_k, D(\alpha, n_k)>$ will be preloaded in each sensor. Also, node ID, gross detection algorithm and other required materials will be preloaded in sensors before deployment. 2) Upon deployment, the cluster head will calculate the transmitting probability P according to the number of inner-cluster nodes and broadcast P to all inner-cluster nodes. Assume that the number of inner-cluster nodes is N and the preloaded gross error parameters are $<n_1, D(\alpha, n_1)>$, $<n_2, D(\alpha, n_2)>$, .., $<n_k, D(\alpha, n_k)>$, the transmitting probability should satisfies: $n_1 \leqslant N*P \leqslant n_k$. For assuring the size of sample set within the bounds of preloaded parameters, the transmitting probability P should higher than $n1/N$. 3) Upon receiving the data request, node Si detects the surrounding parameters and determines if it will transmit its reading ($data_i$) to aggregator according to the transmitting probability P. Assume that the aggregator receives $<data_1, \ldots data_m>$ and sorts the received messages as $< data'_1, \ldots data'_m >$, the statistics r_{ij} and r'_{ij} will be calculated at first, and then, the gross error detection will be performed according to equation (3) and preloaded critical value $D(\alpha, n_i)$. The process result has high reliability and be regarded as the required sample set for data aggregation. 4) The inner-cluster aggregation result will be transmitted from cluster head to the base station. Because main concern of this paper is to reduce the inner-cluster transmissions, thus we don't issue fusion data transmitting in detail.

Simulation

Assume that sensor readings are follow normal distribution with $\mu=30$ and $\sigma=0.1$, take $[-0.01*\mu, 0.01*\mu]$ as acceptable threshhold; the size of sample set ranges from 3 to 15, the significant parameter α is set to 0.01. N, the number of inner-cluster nodes, equals 40; transmitting probability P is set to 0.275 and Pn, the probability of sensor reading is disturbed, equals 0.1. Also, we define the value of disturbed data equals normal data plus 50 and adopt Dixon criterion to detect gross error in PTDA and TCDA for fairness. Also, if the number of incoming sensor readings is greater than the preloaded gross error detection parameter n_k, then only top n_k readings are used for aggregation to ensure the gross error detection can be performed effectively.

The number of inner-cluster transmissions and included abnormal readings of PTDA and TCDA on 50 trails was shown in Fig.2. Where N-TCDA and N-PTDA denote the number of inner-cluster transmissions, NA-TCDA and NA-PTDA denote the number of included abnormal readings.

From Fig.2 we know that the inner-cluster transmissions of PTDA lower than that of TCDA. The average number of inner-cluster transmissions in PTDA is 11, approximately 27.5 percents of TCDA, in accordance with the pre-defined transmitting probability P. Due to the probabilistic transmitting, the number of abnormal data received by aggregator in PTDA lowers than that of TCDA. The average number of received abnormal data in PTDA and TCDA are 1 and 4 respectively, shows the consistency with pre-defined parameter Pn.

Fig.3 (a) shows the inner-cluster aggregation error and Fig.3 (b) shows the aggregation result with Dixon detection. From Fig.3 (a) we know that the abnormal readings may not appear in PTDA (such as trail 18, 19 and 20) because the probability of data transmitting and data disturbing, but the abnormal readings must appear in TCDA due to all nodes will transmit their readings. At the same time, if the abnormal data received by aggregator, PTDA will suffer more from big noise than TCDA. From Fig.3 (b) we know that the aggregation in TCDA is more stable than that of PTDA because the size of sample set of all nodes transmitting larger than that of probabilistic transmitting. Upon gross error detection, aggregation errors of PTDA and TCDA are range from -0.1 to 0.1 and all of them are within the threshold of acceptable error.

Assume the network was divided into 40 clusters and the average number of inner-cluster nodes is 40. Nodes transmit their readings to aggregator with certain probability (TCDA can be considered as special case of PTDA with transmitting probability equals 1). Fig.4 shows the aggregation result of PTDA and TCDA achieved in base station on 50 trails. From Fig.4 we know that the aggregation

error of TCDA and PTDA are ranged from -0.005 to 0.005 and -0.01 to 0.01 respectively. PTDA is suitable for the applications in which the requirement of data accuracy is not so strict. For example, the required accuracy of temperature is 0.1 in general applications. In this case, there are no significant differences among 30.0, 30.01 and 30.005.
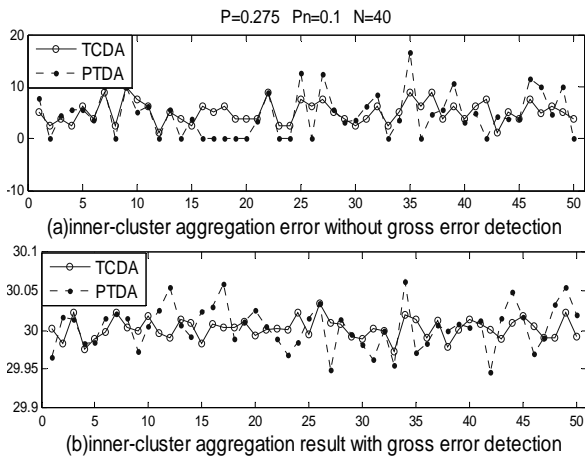

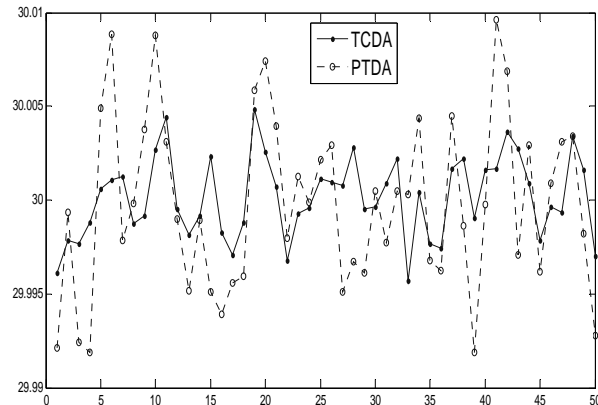Fig.3 the inner-cluster aggregation eror and the inner-cluster aggregation result

Fig.4 the aggregation result in base station

Summary

In this paper, authors proposed a probabilistic transmitting based data aggregation scheme for wireless sensor networks in which each node determines if it transmits data to aggregator according to certain transmitting probability. Experiments show that the proposed scheme lowers the communication overhead of inner-cluster data aggregation effectively and the aggregation results in base station can meet the required accuracy of general applications. This means our scheme is suitable for the energy-limited wireless sensor networks.

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. Araújo, 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.