International Journal of Future Generation Communication and Networking
Vol. 9, No. 1 (2016), pp. 127-134
http://dx.doi.org/10.14257/ijfgcn.2016.9.1.12

Research on Data Aggregation Technology Based on Wireless Sensor Networks

Tianming Wang

Hainan College of Economics and Business,
Haikou 571127, china,
5515180@qq.com

Abstract

The issue of data aggregation in wireless sensor networks was studied. Data aggregation was an efficient in-network data processing method. It reduced data redundancy and improved information quality, which may save communication energy and increase collection efficiency. In this dissertation, data aggregation solution *(DAS) was proposed and realized. DAS used a cluster structure based network and it had three parts. Firstly, the cluster head would not submit any packets derived from the cluster members. Instead, it fused them into an outcome packet and then sent it to the sink node. Secondly, the cluster head scheduled the nodes with low energy level into sleep. Thirdly, the cluster head would use a combined forecasting algorithm to estimate the data of sleeping nodes. Simulation tests were carried out, and simulation results showed that DAS had good performance. It not only extended the network lifetime greatly, but also provided better assurance of data quality.*

Keywords*: Wireless Sensor Networks, Data Collection, Data Transmission, Data Aggregation*

1. Introduction

The imbalance is found between the production and consumption of data in the wireless sensor network [1-3]. On one hand, abundant sensory data are produced every day across the Internet; on the other hand, users concern often too much about the overall parametric indicators of the whole network, instead of the specific reading of some node at one point [4-5]. Hence, it's meaningless to deliver a great deal of data to users, which not only consumes huge energy of network and also reduces users' holistic cognition of monitored targets. Besides, transmission of massive data will lead to channel conflicts and weaken network performance [6-7].

The solution to such contradiction is intranet data processing, that is, in the course of collecting data, perform necessary treatment of them to meet with users' requirements. Data aggregation is one of the main techniques for processing intranet data. By means of contraction and fusion, it fulfills the purpose of removing redundant data and increasing information quality, saving plentiful network energy consumption and boosting collecting efficiency.

Many researchers have probed into such an issue. Earlier studies didn't mention the processing of data semantics, but aimed at the selection of aggregation time and place. They were designed to construct the optimal aggregation routing tree or optimal aggregation cluster structure. The involved fields include routing protocol, clustering algorithm, topology control, compression algorithm and so on, about which, representatives are Tan [8], Krishnamachari [9], Solis [10], and Tang Wei [11]. With the popularization and application of wireless sensor network, there are studies on aggregation mechanism. In the process, data are processed in semantic sense and stronger application relativity is manifested [12]. Common strategies refer to threshold

ISSN: 2233-7857 IJFGCN

suppression, curve fitting, regression analysis, sampling technique etc [13]. They follow the idea: eliminate similarity of low information quantity or repeated data; utilize possibly known data to predict unknown data to decrease submission of data, e.g. use historical data to forecast future data; use partial data to predict whole data etc [14].

Inspired by the above idea, we examine the issue here. We focus on the reduction of data redundancy. It's generally thought that data acquired by sensor nodes have correlations, which have the following two types:

(1) Temporal relativity because one node searches data for a long time at the same location;

(2) Spatial correlation of acquired data as a result of the density of deployed sensor nodes. The existence of such correlation will absolutely give rise to data redundancy. The existing aggregation technologies are mostly investigated from the perspective of temporal relativity, few of spatial relativity. In the paper we consider both kinds of correlation. We propose and implement one data aggregation solution (DAS). The objective of designing DAS is to realize the balance between network energy consumption and information quality. We employ grey model tool to analyze the correlation of submitted data by each node; on that basis, we aggregate data. Moreover, given redundancy of submitted data, we think it's not necessary for all nodes to submit data, letting some nodes sleep. Starting from that point as well, DAS schedules nodes which have lower residual energy to enter into dormant state to cut down node communication overheads and equalize network energy dissipation. The experiment proves well the effectiveness of DAS.

2. Overall Thinking of the Solution

In the network here, after collecting data, nodes would submit them to cluster head nodes. Since temporal and spatial correlation exists in monitored data, such data will have information redundancy. The submission of redundant data results in the waste of node energy and increases conflicts of wireless channels, reducing the quality of communication. DAS algorithm starts from the angle of decreasing information redundancy, making nodes decline unnecessary transmission of data as to lessen network energy consumption. In order to strike balance between data quality and network performance, DAS takes advantage of such technological means as data aggregation, node scheduling, data prediction etc.

2.1 Data Aggregation

In DAS, when the head node of each cluster receives sensory data sent by its members, it doesn't deliver directly to the converging node, but aggregate them and then submit.

2.2 Node Scheduling

One feature of DAS is having introduced node scheduling to data aggregation strategy. In previous studies, the processing of redundant data is to control nodes' sending behavior, for the purpose of saving communication energy. DAS exerts that idea. It lets nodes in dormant state to cut down energy consumption by sending and monitoring. Furthermore, as far as balanced network energy consumption is concerned, DAS always makes dormant nodes with the lowest left energy.

2.3 Data Prediction

For dormant nodes, DAS uses prediction algorithm to estimate data which should be submitted in order to guarantee the quality of submitted data.

The main idea of DAS algorithm is: when the head node is collecting data delivered by each member, it makes relativity analysis of those data to get the related description of

data in temporal and spatial terms. If data of one node is strongly correlated with its previous data or those submitted by other nodes, the data can be regarded as redundant data, because they can be indirectly obtained through historical data or other nodes' data. Therefore, DAS method will schedule such nodes to go into the dormant state to save energy. DAS always schedules nodes with lowest energy level to sleep. Data that deserve submission are induced based on analytics of correlation by the cluster head node.

In the above process, the key step is to judge the correlation of collected data by nodes. It is the foundation for determining data redundancy. In DAS algorithm, correlation judgment includes temporal correlation and spatial correlation. DAS algorithm applies grey model's related theories to respectively process them; at last, with combined prediction model, it aggregates results of them and achieves better effects.

3. Analysis of Temporal Correlation

Temporal correlation analysis refers to analyzing the relevance of one node's acquired data sequence and getting a new data sequence. The new sequence hides the temporal analysis results of that node data by DAS algorithm. It contains the fitted value for existing data and also estimated value for future acquired data. If the node is dormant, the cluster head node can use estimated value to replace its real acquired data.

Temporal correlation analysis is completed with the use of unbiased grey predicative model. The model was an improved model on the basis of traditional Grey model (1,1) by Ji Peirong et al. It computes faster and can be used more widely.

DAS analysis of temporal correlation is carried out in following steps:

(1) Build original sequence

Regard one node for dormancy as examined one; choose consecutive t collected data submitted by the node as original data sequence, i.e. $M^{(0)}$:

$$M^{(0)} = \{M^{(0)}(1), M^{(0)}(2), ..., M^{(0)}(t)\} \qquad (1)$$

(2) Pre-treatment of original sequence

Perform accumulated operation of $M^{(0)}$ to generate a new sequence, as $M^{(1)}$:

$$M^{(1)} = \{M^{(1)}(1), M^{(1)}(2), ..., M^{(1)}(t)\} \qquad (2)$$

(3) Solve GM (1, 1) model

Use $M^{(1)}$ as original input data of GM (1, 1); solve parameter a and μ in differential equation (3);

$$\frac{dM^{(1)}(t)}{dt} + aM^{(1)}(t) = \mu \qquad (3)$$

(4) Solve unbiased GM (1, 1) model

As per a and μ, use expression (4) to solve parameter ϕ and γ in unbiased GM (1, 1) model;

$$\begin{cases} \varphi = \ln \dfrac{2-a}{2+a} \\ \gamma = \dfrac{2\mu}{2+a} \end{cases} \qquad (4)$$

(5) Generate data sequence model

According to ϕ and γ, generate a new data sequence model $\hat{M}$. It is shown in expression (5).

$$\hat{M}(i) = \begin{cases} M^{(0)}(1) & i = 1 \\ \gamma e^{\varphi(i-1)} & i > 1 \end{cases} \quad (5)$$

4. Flow of DAS Algorithm

DAS is implemented as follows:

(1) Network initialization and building of cluster structure
(2) Set parameters like dormant rate r and dormant cycle timer etc;
(3) Network starts working: the cluster head node collects data of all its members; to guarantee the precision of subsequent analysis of data correlation, in the initial running time t range of network, all nodes need to submit data; when submitting data, nodes notify the cluster head node of their own left energy; the cluster head performs fusion processing after receiving node data and delivers to the converging node;
(4) From its own cluster, the cluster head node finds out some nodes with the lowest energy according to the dormancy ratio r and informs them to sleep; the dormant cycle timer is simultaneously initiated;
(5) Non-dormant nodes submit data: for dormant nodes, the cluster head node makes temporal and spatial relativity analysis based on submitted data and uses combined predictive method to estimate data which should be delivered during the dormancy; on that basis, the cluster head node integrates all data and submit;
(6) Once the dormant cycle counter is time-up, the cluster head node will awake dormant nodes to work again; go back to (4).

5. Experiment Design and Discussion

5.1. Testing Scene and Indicators

In the paper, we realized DAS protocol in NS2 environment. To validate its performance, we implemented several other protocols and algorithms; then we made comparative tests between them.

Several other protocols are:
(1)Common data aggregation scheme (Common)
(2)Randomized Independent Sleeping based data aggregation scheme (RIS)
(3)Least residual Energy node sleeping based data aggregation scheme (LE)
(4) Temporal correlation based Data Aggregation scheme (TDA)
(5) Spatial correlation based Data Aggregation scheme (SDA)

Of the above five protocols, Common protocol fuses only data which are submitted by cluster members to their cluster head node, without any analysis of node scheduling and correlation [15]; RIS protocol introduces on the basis of Common the random independent dormant algorithm, which means during each cycle, nodes enter into dormant state at a certain probability; RIS protocol doesn't make correlation analysis and estimation of data of dormant nodes; LE protocol is also evolved based on Common; it adopts the dormant strategy of dispatching nodes with the lowest residual energy to dormant state and doesn't estimate data of dormant nodes; being the improvement of LE, TDA analyzes temporal correlation of data of dormant nodes and estimates what data should be submitted; SDA develops on the basis of LE protocol; what's different is SDA analyzes spatial correlation of data of dormant nodes. From that, we learn that compared with DAS protocol, TDA and SDA stand respectively for the situation when $\rho=1$ and $\rho=0$.

The setting of the experimental parameters is shown in table1.

The testing data here were collected from GreenOrbs project [16]. Figure 1 shows nodes used by GreenOrbs and the scope of deployment [17]. GreenOrbs is a long running forest ecology monitoring system. It can collect lots of information like temperature, humidity, light intensity, concentration of carbon dioxide. It's mainly used for estimating canopy closure, studying ecological diversity, carbon sequestration, fire monitoring etc.

This paper mainly used GreenOrbs temperature and humidity data.

In the paper we compare performance of those protocols from the part of information quality and network energy consumption, with the two indices:

(1)Network life cycle: the time when the first node dies in the network; it measures the level of network energy consumption;

(2)Data mean error: the error mean between the actual submitted data by the cluster head node and theoretical data; theoretical data means fusion data submitted by cluster nodes when they're working normally.

Table 1. Experimental Parameter Settings

Parameter	Values
Node number	81
Routing strategy	Ordinary node: directly sent to the cluster head node Cluster head node: data is sent directly to the sink node.
MAC protocol	802.11
Node initial energy	3J
Sink node energy	10J
Cluster head node energy	10J
Array packet size	2Byte
Network cluster number	Three
Data sending cycle	1 times per second
Cluster head fusion scheme	The average value of the data in the cluster

Figure 1.Greenorbs Node and Deployment Range

5.2. Simulation Results and Discussion

5.2.1. Network Life Cycle

Comparison of network life cycle is shown in Figure 2. The diagram uses the protocol of node scheduling strategy and that network life cycle is all prolonged. Common protocol doesn't dispatch nodes to sleep and thus the testing result tends to be a horizontal line. Lines of TDA, SDA and DAS overlap LE's because they adopt the scheduling strategy that nodes with lower energy should firstly become dormant; moreover, they are based on similar network topological structure with the same communication model; that's why the testing result is identical. Compared with RIS protocol, the three protocols work much better, because RIS randomly assigns nodes to sleep, which may lead to uneven energy consumption by nodes, and thus the life cycle becomes shorter.

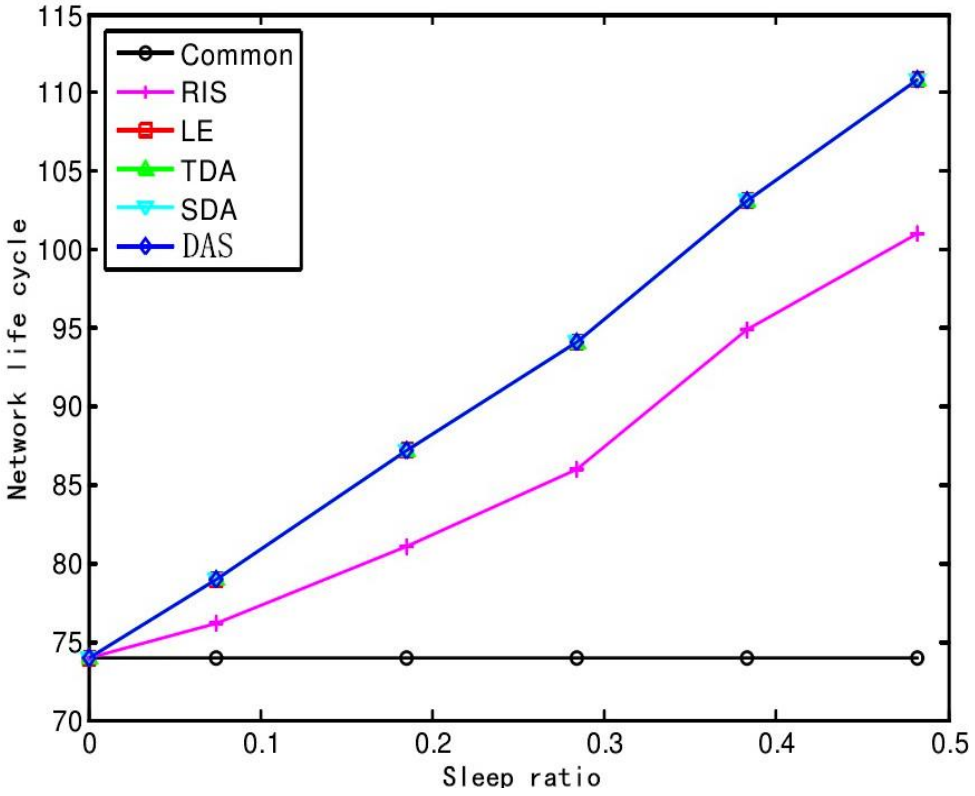

Figure 2. Network Life Cycle Comparison

5.2.2. Mean Error of Temperature Data

Temperature data errors are compared in Figure 3. In it, the vertical axis stands for the ratio between temperature error and theoretical data. From the picture, we see Common protocol collects all data and therefore no error is found for testing result. Errors of other protocols become ascending with higher percentage of dormant nodes. RIS protocol and LE protocol both have higher errors than TDA, SDA and DAS because they don't use any estimation method to make up data of dormant nodes. Of three protocols which use correlation prediction, DAS has the lowest error since it combines the result of both temporal and spatial correlation analysis.

5.2.3. Average Error of Humidity Data

Figure 4 compares the error of temperature data, where axis Y is the ratio between humidity error and theoretical data. Due to big variation of humidity data, the overall

error is raised. As far as testing result is concerned, like testing results regarding temperature, except Common, all other protocols' errors go higher with more dormant nodes. TDA, SDA and DAS use certain estimation technology. That's why their errors increase smaller than RIS and LE. Besides, DAS combines the result of both temporal and spatial correlation analysis. That's why its error is the lowest.

What's more from Figures 3-4, we observe that protocols making spatial correlation analysis have lower errors than those making temporal correlation analysis, i.e. SDA performs better than TDA. It indicates that in the testing data, the spatial correlation of collected data is stronger than temporal correlation, which gives rise to better testing effects.

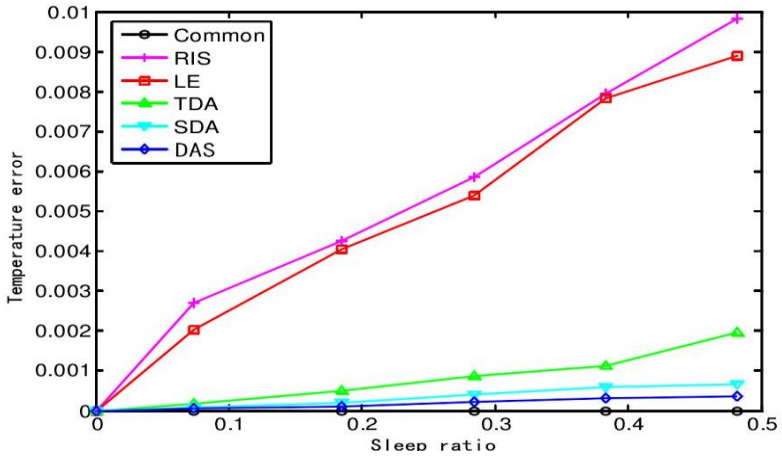

Figure 3. Temperature Error Comparison

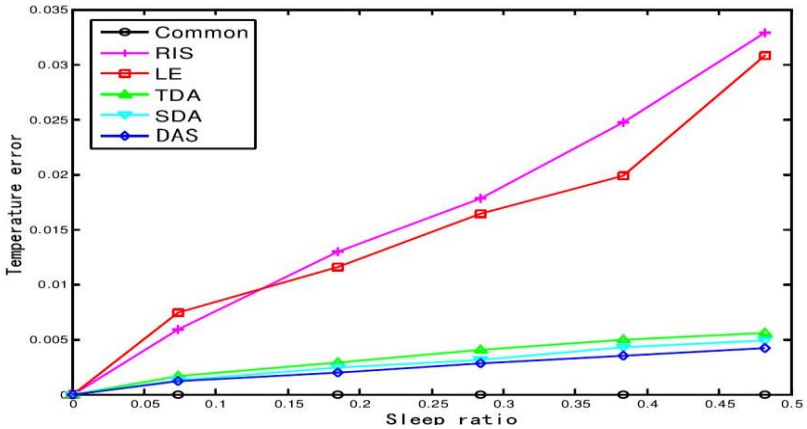

Figure 4. Temperature Error Comparison

6. Conclusion

We discussed the aggregation problem of collected data. In the wireless sensor network, users concern mostly about parametric indicators of the whole network, instead of the specific reading of some node at one point. In this case, transferring tremendous data to users not only becomes insignificant and also causes more channel conflicts and more energy consumption by nodes. To overcome that, from the perspective of energy consumption, we designed and implemented a data aggregation solution based on temporal and space correlation, which is shortly DAS. DAS uses cluster structure network. It includes three key points:

(1) The cluster head node fuses data inside the cluster and then delivers to the sink node;

(2) The cluster head node schedules nodes with the lowest energy to become dormant;

(3) The cluster head node utilizes combined prediction algorithm to estimate data of dormant nodes. Simulation testing revealed that DAS made better effect. The proposed method is not only an effective data aggregation approach but also a better node scheduling method.

References

[1] C. Xuehan, C. Feng and W. Jia, "A WSN based on the spatio-temporal correlation of the network data aggregation routing protocol", Computer engineering and science, vol. 01, **(2015)**, pp. 48-55.
[2] F. Huiying and D. Qiulin, "Data aggregation algorithm for wireless sensor networks based on data stream and network encoding", Computer science, vol. 05, **(2015)**, pp. 136-141.
[3] T. Wei and G. Wei, "The maximum lifetime gene routing algorithm in wireless sensor networks", Journal of software, vol. 07, **(2010)**, pp. 1646-1656.
[4] Z. Qiang, L. Xiao and X.C. Cui, "Based on clustering wireless sensor network data aggregation scheme research", Journal of sensor technology, vol. 12, **(2010)**, pp. 1778-1782.
[5] L. Hong, "H. Study on polymerization technology of wireless sensor network data support QoS", Application Research of computers, vol. 01, **(2008)**, pp. 64-67.
[6] W. Ruchuan, "A method based on the estimation of the cost of data aggregation tree generation algorithm in sensor networks", Electronic journal, vol. 05**, (2007)**, pp. 806-810.
[7] W. Zhu, W. Debao and W. Ling, "The timing control algorithm for data aggregation in sensor networks", Instrument technique and sensor, vol. 05, **(2012)**, pp. 75-78.
[8] H.O. Tan and I. Körpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks ", SIGMOD Record, vol. 32, no. 4, **(2003)**, pp. 66-71.
[9] B. Krishnamachari, D. Estrin and S. Wicker, "Modelling Data-Centric Routing in Wireless Sensor Networks", Proceedings of the International Conference on Computer Communications, **(2002)**, pp. 42-49.
[10] I. Solis and K. Obraczka, "The impact of timing in data aggregation for sensor networks", Proceedings of the IEEE International Conference on Communications, **(2004)**, pp. 3640-3645.
[11] T. Wei and G. Wei, "Dalkey station data aggregation in wireless sensor network maximum lifetime geographic location routing", Journal of communication, vol. 31, no. 10, **(2010)**, pp. 221-228
[12] Z. Deyun, Y. Jun and Z. Yunyi, "Data aggregation and transmission protocol of wireless sensor networks based on clustering", vol. 21, no. 5, **(2010)**, pp. 1127-1137.
[13] C.Q. Zhang, M.L. Li and M.Y. Wu, "A model aided data gathering appr oach for wireless sensor networks", IEEE Transactions on Wireless Communications, vol. 22, no. 4, **(2007)**, pp. 27-30.
[14] Y. Kotidis, "Snapshot queries: towards data-centric sensor networks", Proceedings of the International Conference on Data Engineering, **(2012)**, pp. 131-142.
[15] S. Kumar, T.H. Lai and J. Balogh, "On k-coverage in a mostly sleeping sensor network", Wireless Networks, vol. 14, no. 3, **(2008)**, pp. 277-294.
[16] Y.H. Liu, Y. He and M. Le, "Does Wireless Sensor Network Scale? A Measurement Study on GreenOrbs", Proceedings of the International Conference on Computer Communications, **(2011)**, pp. 873-881.
[17] L.F. Mo, Y. He and Y.H. Liu, "Canopy Closure Estimates with GreenOrbs: Sustainable Sensing in the Forest", Proceedings of the ACM Conference on Embedded Networked Sensor Systems, **(2009)**, pp. 99-112.

Author

Tianming Wang, He received his M.S degree from Changchun University of Technology. He is a lecturer in Hainan College of Economics and Business. His research interests include Computer network.

