http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology of Journal 4 (4): 462-464, 2005
ISSN 1812-5638

Data Aggregation with Error Correction for Wireless Sensor Networks

Qiang Ma and Won-Sik Yoon
Department of Electrical and Computer Engineering, Ajou University, Suwon, Korea

Abstract: In-network aggregation as a power-efficient mechanism for collecting data in wireless sensor networks has been emphasized with the development of other WSN network protocols. In this study, we firstly propose a general architecture to seamlessly integrate data aggregation into wireless sensor networks and then present one error correction algorithm for data aggregation by exploiting the inherent correlations that exist between sensor nodes. This approach is orthogonal to other works and thus can work complementally to enhance the reliability of aggregation algorithms

Key words: Fault tolerant, data aggregation, wireless sensor networks

INTRODUCTION

A typical task of WSN is the monitoring of a large area with respect to some given physical parameters, e.g. temperature and humidity. The information is gathered by sensors and reported to one or several points which are usually referred as data sinks/requesters. A possibility to reduce the amount of data to be transmitted and therefore to conserve energy, is to combine several sensor readings in intermediate nodes along the way towards the sinks. This process is referred to as in-network aggregation. Data aggregation procedure usually consists of distribution phase and collection phase[1]. The sink node firstly requests readings from the entire network by flooding the network with appropriate queries[2] and accordingly a convergecast tree, along which the answers are reported back to the sink, is constructed.

Timing synchronization[3] is an important issue in data aggregation. Each parent waits for children's readings before it aggregates with its own readings. Timing model defines when to clock out data. There are typically three timing models: periodic simple aggregation in which each node waits for a pre-defined period of time, aggregates all data items received and then sends out a single packet; periodic per-hop aggregation which transmits the aggregated data as soon as it hears from all its children; periodic per-hop adjusted which adjusts timeout based on their position in the data collection tree. Timing models may have significant impact on the freshness and accuracy of data delivered by aggregation algorithms[4]. Upadhyayula *et al.*[5] proposed a heuristic solution for the problem of minimum energy convergecast which also works toward minimization data latency[5]. Mukhopadhyay *et al.*[6] proposed an error-prone algorithm by exploring temporal correlation.

In this study fault tolerant data aggregation has been proposed to enhance the reliability of data aggregation algorithms by exploring spatial correlation existing in sensor data. The proposed method is orthogonal to other approaches and thus can be used in conjunction with some of them. Also the proposed method has the advantages such as energy efficiency, no delay induced and no dependence on the underlying layers.

BACKGROUND AND MOTIVATION

The main goal of data aggregation is to minimize the total transmission cost of transporting the information collected by the nodes, to the sink node. Correlations typically exist among sensors that are measuring data in the same geographical location[7]. For example, temperature sensors or humidity sensors in a similar geographic region produce readings that are correlated. Another example is audio sensors that are sensing a common event such as a concert or whale cries will produce measurements that are correlated. In most case sensor readings will not change dramatically during a short time period and can be predicted in a short time range.

Due to the dynamics of wireless ad hoc network, transmissions are error-prone. Messages may not be received correctly and thus the aggregation result may be totally wrong. For example, a typical aggregation function is minimum, maximum or average. If the minimum reading or maximum reading is wrong or lost, the aggregation result is therefore wrong. In general, the transmission cost of sending 1 bit of data costs as much as executing 1000 CPU instructions[8]. Thus conventional recovery mechanisms such as ARQ or FEC are not suitable in wireless sensor networks because those techniques either waste the precious energy to transmit additional bits or induce delay which accordingly complicates the time synchronization algorithm.

Corresponding Author: Won-Sik Yoon, Telecom Research Lab, Department of Electrical and Computer Engineering, Ajou University, 5 Woncheon-Dong, Youngtong-Gu, Suwon, 443-749, Korea
Tel: +82-31-219-2371 Fax: +82-31-212-9531

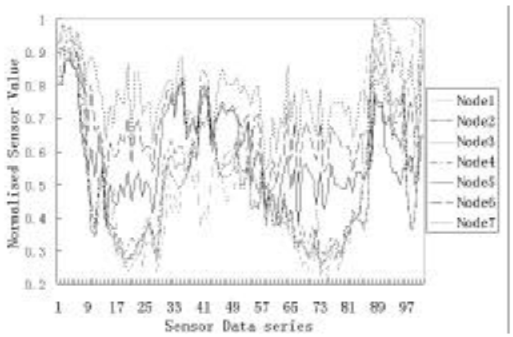


Fig. 1: Spatial correlation among sensor readings

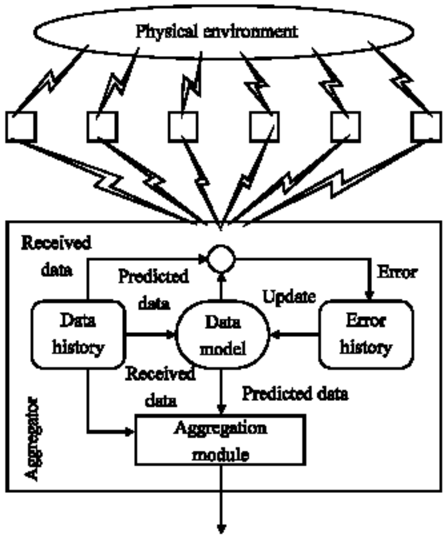


Fig. 2: Data aggregation framework

Some sampled readings of the node do not need to be processed and transmitted to its parent when its future values can be approximately estimated from the prediction function, which leads to an end-to-end approach that does not require communications between sensor nodes. Figure 1 shows plots of 100 speed data values collected by sensors in ITS project[9]. It can be seen that there is a strong spatial correlation between nodes.

DATA AGGREGATION FRAMEWORK

This newly designed method answers the following two questions: how to obtain such a prediction model and how to guarantee the difference between the predicted and the real value. Figure 2 describes the main ideas and methods to seamlessly integrate prediction techniques into data aggregation framework. Within each aggregator a data model is constructed on-line based on the correlation observed in sensor data; this model is used during data acquisition for online data prediction; a running history is maintained of the observed values as well as prediction errors. These histories are then used together with the data model to generate predictions and possibly update the model as well.

The operation of the aggregation module block is independent of the data model used for prediction. A variety of modeling techniques can thus be used to represent data correlation properties. However, the performance of the correction algorithm depends on the accuracy of modeling and efficiency of the prediction.

DATA MODELING AND EXPERIMENT

The applicability of this approach relies upon representative samples of the sensor data being available in order to build accurate predictive models. This includes all applications that observe the current state of some physical process and monitor them for known or unknown variations.

The pseudo-code of the algorithm used to correct errors at the aggregator is shown in Fig. 3. r is the control parameter-prediction error threshold. Once r is beyond the threshold, update CC function will be called to select another data set and update data model parameters. Another numeric variable s is used to record the selected data set. The history of observed and predicted values and the corresponding prediction errors is stored in the data history and error history lists. A low prediction error threshold in each aggregation

begin
for each period some sensor readings are missing or wrong, calculate the regression coefficient a and b

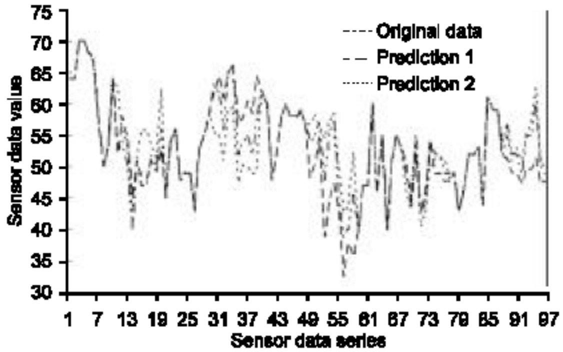
$$b = \frac{cov_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}$$

Where $\bar{x}$ and $\bar{y}$ are the mean of data set x_i and y_i, s_x is the standard deviations of xi and cov_{xy} is the covariance
use the predicted value to replace the wrong or lost data end
update CC (y_i, x_i, r)
for each completed data set, calculate the correlation coefficient r with the uncomplete data set;

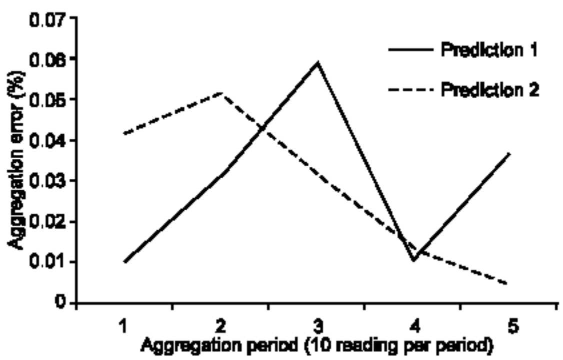
$$r = \frac{cov_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2)^{1/2}}$$

find $r = r_{max}$, which shows strong correlation;
end

Fig. 3: Prediction algorithm pseudo code



(a) Data prediction



(b) Aggregation accuracy

Fig. 4: Prediction based on spatial correlation

period can be chose to reduce the difference between the real value and predicted one or set a fixed prediction error threshold and update the data model parameter if the prediction error is beyond tolerance.

In this scenario, ten readings from each node are aggregated each period using average function at the aggregator. And aggregation error is represented by the root mean square sum of residual errors of the aggregated sensor readings. It is assumed that data set from node 5 is incomplete and is marked as original data in Fig. 3. This algorithm used with the best two correlated data sets-data readings collected by node 1 and node 2-for prediction. Figure 4a shows the compared results with the original data. Prediction 1 is based on data set from node 1 and prediction 2 is based on data set from node 2. Figure 4b shows that this method achieves a good performance by limiting the aggregation error within 1%.

CONCLUSIONS

In this study, an error correction algorithm has been proposed for data aggregation. This approach is orthogonal to other data aggregation algorithms and thus can work complementally to achieve better performance. In addition to exploiting spatial correlation, the temporal correlation properties of sensor data can also be used to do error correction for data aggregation.

REFERENCES

1. Madden, S. R., M. J. Franklin, J. M. Hellerstein and W. Hong, 2002. TAG: A tiny aggregation service for ad-hoc sensor networks. In Operating System Design and Implementation (OSDI).
2. Intanagonwiwat, C., R. Govindan, D. Estrin, J. Heidemann and F. Silva, 2003. Directed diffusion for wireless sensor networking. IEEE/ACM Trans. Networking, 11: 2-16.
3. Sivrikaya, F. and B. Yener, 2004. Time synchronization in sensor networks: A survey. IEEE Network, 18: 45-50.
4. Solis, I. and K. OBraczka, 2004. The impact of timing in data aggregation for sensor networks. IEEE Intl. Conf. on Communi., 6: 3640-3645.
5. Upadhyayula, S., V. Annamalai and S.K.S Gupta, 2003. A low-latency and energy-efficient algorithm for convergecast in wireless sensor networks. Global Telecommuni. Conf., 6: 3525-3530.
6. Mukhopadhyay, S., D. Panigrahi and S. Dey, 2004. Data aware, low cost error correction for wireless sensor networks. Wireless Communications and Networking Conference, pp: 2492-2497.
7. Pattem, S., B. Krishnamachari and R. Govindan, 2004. The impact of spatial correlation on routing with compression in wireless sensor networks. Proceedings of Information Processing in Sensor Networks, Berkeley, CA, pp: 28-35.
8. Emekci, F., H. Yu, D. Agrawal and A.E. Abbadi, 2003. Energy-conscious data aggregation over large-scale sensor networks. Technical Report, Computer Science, University of California Sant Barbara.
9. http://trl.ajou.ac.kr Apr., 2005