IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Survey on Data Aggregation Techniques

Prasad B. Thube[1], Sachin B. Takale[2]

PG Student, Dept of Electronics and Telecommunication Engineering, Sinhgad College Engineering, Pune, India [1]

Professor, Dept of Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Pune, India [2]

Abstract: The wireless sensor node have much application such as forest monitoring, habitat monitoring, security and military, disaster management, etc. Wireless sensor nodes are resource constrained, sensor nodes are small in size and have limited memory, low battery power and limited processing capability. Data aggregation is very important technique is wireless sensor node. Data aggregation eliminates the redundancy data helps to reduce to requirement of store the data and reduce power consumption. This works shows different data aggregation technique and comparative analysis between aggregation techniques.

Keywords: Wireless Sensor node, Data Aggregation, spatial correlation, time aggregation.

I. INTRODUCTION

Data aggregation is the method to collect data and aggregate the useful data. Data aggregation is considered as fundamental procedure for saving memory. In wireless sensor node, data aggregation is an effective way to save the limited resources. The main goal of the data aggregation method is to gather and aggregate data in a memory efficient manner so that increase the node lifetime is increase. Wireless sensor node has limited memory, limited computational power and battery power due to this increase the complexity for application developers. In this paper, a data aggregation technique for wireless sensor node is presented along with survey on data aggregation algorithms

(i) An Overview on Data Aggregation

The data aggregation is a method used to solve the similar data problems. Data source coming from sensor node is aggregated if they are similar data. Data aggregation is a widely used method in wireless sensor node. The data confidentiality and integrity security issues, in data aggregation become more vital when the sensor node is deployed in a harsh environment. Data aggregation is a process of the aggregating sensor information using aggregation protocol.

(ii) The Need of Data Aggregation

Sensor nodes are deployed in remote environment in wide range of area. Very rarely the user has useful information on the sensor node. Due to this when sensor reading of the attributes like temperature, pressure humidity in an arbitrary area, node may suffer the heavy traffic. This problem needs data aggregation to act with user requirements and manage homogenous data with efficiently. Many practical applications such as environment monitoring, scientific research etc., are exploring the use of wireless sensor nodes. Such application transferring a huge amount of data sensed data. Since wireless sensor nodes are with low power batteries, battery life is a major key role in real time application. This required data dissemination for power and memory efficient protocol for aggregation of sense data. Node in close proximity usually holds homogenous data due to a property called spatial correlation. Memory and power is wasted when the homogeneous data value is processed. It is desirable to process as much as data at node level to reduce the number of bits transmitted through air, for long distance. It is an also known that, to transmitting 1KB over a 100 m [4] costs the similar amount of power as executing 300 million instruction on a general purpose processor with the modern computing device rate of 100 million instruction per second (MIPS). In data aggregation protocol ideally, each senor node spending the equal amount of power in each data gathering in close.

A data aggregation scheme is memory and power efficient if it maximizes the functionality of sensor node. We assume if all sensors are equally important, we should minimize memory requirement and power consumption of sensor node. The first step is data collection from sources and aggregation of data. [6]

II. DATA AGGREGATION TECHNIQUES

There is various type of data aggregation

- RMS
- Sum and Count
- Min and Max

a) RMS

The RMS defined as the square root of the mean of square of samples. This aggregation protocol related to voltage measurement is the root mean square algorithm act on quantities a like energy transfer. The RMS aggregation algorithm uses 4 different time interval: 200ms, 3s, 10 minutes, 2 hours as aggregation interval. RMS aggregation performed using the square root of arith mean of squared input sample. Typically there are three categories of aggregation are done: cycle aggregation, package aggregation and time-clock aggregation [1]. A compromise between the user's needs for data and amount of information resulting from averaging on one period of

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

observed quantities waveform i.e. RMS aggregation has been up to now done on the time axis. Figure 1 shows the theoretical structure of RMS aggregation process. The sensed data aggregation is given by the use of RMS values. Below figure shows the RMS aggregation algorithm waeform.

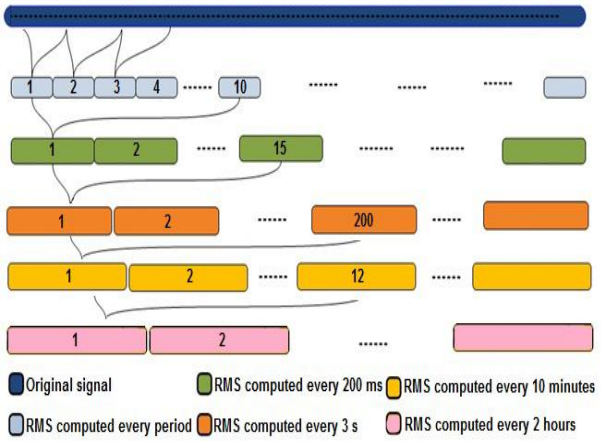

Fig 1: RMS aggregations process [6]

The 2 hours interval aggregated data should be computed from twelve 10 minute time interval. The formula for the first level (0.2s time interval) is

$$v_{(0.2s)} = \sqrt{\sum_{i=1}^{10} \frac{v_i^2}{10}} \qquad (1)$$

Where v_i is the quantity associated with one fundamental period of sinusoidal waveforms. This sample is the RMS value of voltage or current quantity. The parameter like man and variance should not change according to time. The formula used for further three time aggregation interval sample (3 s, 10 minute and 2 hours) are same as formula (1) only changes in time interval. The sense of data compression is given by the use of RMS sample values. [2]

b) Count and Sum

The important aggregation algorithm considered by research community is count and sum. To generalize this aggregation to predicts count and sum. Furthermore, Average can be calculated from the count and sum algorithm. A sum protocol can be extending to compute standard deviation of any order. Count function sensitive to duplicate. The average calculated from count and sum. The formula for average is

$$\text{Average} = \text{Sum} / \text{Count} \qquad (2)$$

c) Minimum and Maximum

Depending on the application the data of sensor node aggregating into minimum and maximum values. The select value of minimum and maximum from different values of sensor node. This methodology provides a fault-tolerant solution. Simple computation can easily deliver duplicate data to a node. A min or max using an iterative computation results in tree data. The min and max can be computed using the function call digest diffusion. For example, suppose we are interested in calculating a digest that is representing the value of minimum energy at node call this value E min. [3]

III. IMPLEMENTATION ISSUES

Here, we analysed the performance issues of data aggregation. Few researchers' algorithms which attempts to solve more difficult problem than aggregation at the cost of more communication overhead and latency. We report the performance issues of data aggregation.

i) Latency

Protocol completes within one epoch simultaneously with the original algorithm. The worst case latency includes in the sum and count is O (log S /W), where S is the upper bound of sum and W is the size of sliding window used. Note that if the upper bound of sum is large, then can incur high latency.

ii) Communication Overhead

Note that the sum grows with the number of modes as well as sensed value. Recently, the researchers proposed a modification that reduces the communication per node to O (Δ log N), where Δ is the number of node. The computational overhead is evaluated in terms of computation required for data aggregation. The number of computation in the node is function of percentage of data redundancy. Percentage of data redundancy is defined as being ratio of the redundant data to that of total generated data by sensor nodes. The data redundancy is included to show the benefit of data aggregation in densely deployed sensor nodes.

iii) Approximation Error in data aggregation

The current algorithm, the algorithm produces an approximate estimate of the aggregate, where the amount of error is reduced if the number of synopses used is increased. Algorithms return the exact estimate if no message is lost. The algorithms produce an approximate estimate. Performance comparison of various techniques of data aggregation shown in table1 the performance analysis of the various data aggregation technique

TABLE 1: Performance analysis of the data aggregation technique

Characteristics	RMS	Sum and Count	Minimum and Maximum
Latency (ephos)	Yes	Yes	No
Approx. error (byte)	Yes	Yes	Yes
Overhead (bit)	Yes	Yes	No
Computational complexity	High	Medium	Less
Accuracy	High	Medium	Less

DOI 10.17148/IJARCCE.2016.55160

IV. CONCLUSION

Data aggregation is an important primitive for sensor operation. The aggregation algorithms to compute aggregates such as predicate count and sum. The amount of error is estimate of aggregates. The successful computation of aggregation in the presence of the error. The aggregation leads to low data availability. The flexibly achieve the balance between memory, timeliness and data availability. The aim is to reduce memory requirement to store data. The performance analysis indicates that the computational and communication overhead is not substantial. The simulation results show that the percentage of transmitted data is reduced by up to 60%, leading to a significant improvement in bandwidth utilization and energy consumption. As for future research, we consider of enabling every sensor node to be capable of both aggregating and sending data in order to improve node efficiency.

ACKNOLWEDGEMENT

Authors would like to express their gratitude towards respected PG coordinator **Prof. M. M. Jadhav** and Head of Department, **Dr. M. B. Mali,** for their dynamic and valuable guidance in fulfilment of paper. Authors would also like to thank the principal **DR. S. D. Lokhande,** SCOE Pune, for providing with all necessary infrastructure and facilities to complete the survey paper.

REFERENCES

[1] M. Albu, “dual data aggregation for power quality assessment”, instrumentation and measurement technology conference (I2MTC) proceedings, 2014 IEEE international, MAY 2014.

[2] Ana Maria Dumitrescu, Mihai Calin, Mihaela M. Albu, “Measurement Data Aggregation in Power Systems – Challenges and Opportunities”, Proc. of the 2nd Conference on the Power Options for the Eastern Mediterraneanregion, POEM, Nicosia., 6-8 October 2013.

[3] Sankardas Roy, Mauro Conti, Sanjeev Setia , sushil Jajodia, “Secure data Aggregation in wireless Sensor Network ”, IEEE transaction on Information forensic and security, 2013.

[4] L. Zhong, R. Shah, C. Guo, and J. Rabaey, “An ultra –low power and distributed access protocol for broadband wireless sensor networks,” presented at the Net world + Interop: IEEE Broadband Wireless Summit, Las Vegas, NV, May 2001.

[5] IEC 61000-4-30 ed2.0, Electromagnetic compatibility (EMC) Part 4-30: Testing and measurement techniques - Power quality measurement methods,2008

[6] A. Tripathi, S. Gupta2, B. Chourasiya, “Survey on Data Aggregation Techniques for Wireless Sensor Networks” International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014

DOI 10.17148/IJARCCE.2016.55160