Study on Data Representation and Aggregation in WSN

M. Lakshmi
Asst Professor
Dept of Comp science
TheM.D.T Hindu College
Tirunelveli

P. Velmani, PhD
Asst Professor
Dept of Comp science
TheM.D.T Hindu College
Tirunelveli

P. Arockiya Jansi Rani, PhD
Associate Professor
Dept of Compsci&Engg
M.S.University
Tirunelveli

ABSTRACT
Wireless sensor networks (WSN) consist of sensor nodes. These sensor nodes have the ability of sensing variety of environment. WSN are very small in size and have limited processing capability and very low battery power. So most of the research is based on to reduce power consumption. One such research direction is data aggregation. Even though energy consumption is an important factor in WSN, sometimes it is necessary to record certain data to predict the future. To do effective data aggregation, data representation is a key point to be considered. The aim of this paper is to study various data representation and data aggregation models to report meaningful data that may be useful for future analysis.

Keywords
Wireless Sensor Network, Pollution Monitoring, Data aggregation, Protocols.

1. INTRODUCTION
WSN has huge application in habitat monitoring, disaster management, security and military, etc. Sensor nodes have the ability of sensing the temperature, pressure, vibration, motion, humidity, sound etc. These sensor nodes consist four main units they are sensing unit, processing unit, transmission unit, power unit. Wireless sensor nodes are very small in size and have limited processing capability with low battery power [1]. The bulk of the research on WSNs have focused on the effective support of the functional, such as data latency, and the nonfunctional, such as data integrity requirements while coping with the resource constraints and on the conservation of the available energy in order to prolong the life of the network[2]. Data aggregation is very crucial technique in WSN. With the help of data aggregation it is possible to reduce the energy consumption by eliminating redundancy[3]. To do effective data aggregation, data representation is a key point to be considered. This paper organized as follows. Section 2 outlines basic ingredients of data aggregation, Section 3 focuses an application scenario of WSN, Section 4 implements some of the models of data representation, data aggregation and their drawbacks and finally section 5 discusses the efficient model by taking into the factor that the aggregation model that may be useful for future analysis and section 6 concludes the paper.

2. BASIC INGREDIENTS OF DATA AGGREGATION
To do effective data aggregation, the important things are to be considered as data representation, data aggregation, routing protocols.

2.1 Data representation
The sensor nodes have limited storage capabilities therefore all the received and generated information cannot be stored. The node has to decide whether to store, discard, compress, or to transmit the data. For performing this task the data has to be represented in an appropriate format[4]. The format varies based on the application requirement. Generally the data structure is common to all nodes, but the main constrain is it should be adaptable to node specific characteristics. The design of a WSN depends significantly on the application and it must consider factors such as the environment, the applications design objectives, cost, hardware and system constraints. .A recent and promising method to deal with data representation and compression is distributed source coding techniques that compress data on the basis of some knowledge about its correlation[5].

2.2 Data aggregation
Data aggregation is the process of collecting and aggregating the useful data. Data aggregation is considered as one of the fundamental processing procedure for saving the energy. In WSN, data aggregation is an effective way to save the limited resources. The main goal of data aggregation algorithm is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced[6]. Data aggregation in WSN can be done by using aggregation function such as MAX, MIN, AVG, COUNT, SUM etc on aggregator node. There are several types of aggregation functions and most of them are closely related to the specific sensor application. Data aggregation is also include eliminating redundant data from various sensor node[7]. Each aggregation functions has some drawbacks during analysis phase. In-network aggregation with size reduction refers to the process of combining and compressing data coming from different sources in order to reduce the information to be sent over the network. In-network aggregation without size reduction refers to the process of merging packets coming from different sources into the same packet without data processing. Timing strategies are required especially in the case of monitoring applications where sensor nodes periodically report their reading to the sink.

Other approaches are lossy Approach-The original values cannot be recovered after having merged them by means of the aggregation function .In addition we may lose in precision with respect to transmitting all readings uncompressed. Lossless Approach-Lossless allows to compress the data by preserving the original information. This means that all reading can be perfectly reconstructed from their aggregate at the receiver side.

Duplicate sensitive and duplicate insensitive-An intermediate node may receive multiple copies of the same information. In this case it may happen that the same data is considered multiple times when the information is aggregated .If the aggregation function in use is duplicate sensitive, the final result depends on the number of times the same value has been considered. Otherwise the aggregation function is said to be duplicate insensitive. For instance, a function that takes the average is duplicate sensitive, where a function that takes the

minimum value is duplicates insensitive.[5] Table 1 shows various data representation and in-network aggregation functions.

2.3 Routing protocols

Routing in sensor networks is very challenging due to several characteristics that distinguish them from contemporary communication and wireless ad hoc networks. First of all, it is not possible to build a global addressing scheme for the deployment of sheer number of sensor nodes. Therefore, classical IP-based protocols cannot be applied to sensor networks. Secondly, in contrary to typical communication networks almost all applications of sensor networks require the flow of sensed data from multiple regions (sources) to a particular sink. Third, generated data traffic has significant redundancy in it since multiple sensors may generate same data within the vicinity of a phenomenon. Such redundancy needs to be exploited by the routing protocols to improve energy and bandwidth utilization. Fourth, sensor nodes are tightly constrained in terms of transmission power, on-board energy, processing capacity and storage and thus require careful resource management. Due to such differences, many new algorithms have been proposed for the problem of routing data in sensor networks. These routing mechanisms have considered the characteristics of sensor nodes along with the application and architecture requirements. Almost all of the routing protocols can be classified as data-centric, hierarchical or location-based although there are few distinct ones based on network flow or quality of service (QoS) awareness[8]. Figure 1 shows the Networking protocols and hierarchies for in-network aggregation

3. AN APPLICATION SCENARIO OF WSN

One of the world wide application of WSN ,that is need of the hour is pollution monitoring. A Comprehensive Environmental Pollution Index(CEPI),which is a rational number to characterize the environmental quality at a given location. The main objective of the study is to identify polluted industrial clusters or areas in order to take action and to centrally monitor them to improve the current status of their environment components such as air and water quality data, ecological damage, and visual environment conditions.

Air pollution due to industries and increased traffic is becoming an increasingly dominant problem, especially in the developing countries. It is observed that the levels of one or more air pollutants are higher than the permissible limits set by Central Pollution Control Board in many industrialized/urbanized pockets of India.

The Clean Air Act requires every state is to establish a network of air monitoring stations for criteria pollutants, using criteria set by OAQPS (Office of Air Quality Planning and Standards) for their location and operation. The monitoring stations in this network are called the State and Local Air Monitoring Stations (SLAMS). The states must provide OAQPS with an annual summary of monitoring results at each SLAMS monitor, and detailed results must be available to OAQPS upon request. To obtain more timely and detailed information about air quality in strategic locations across the nation, OAQPS established an additional network of monitors: the National Air Monitoring Stations (NAMS). NAMS sites, are part of the SLAMS network, must meet more stringent monitor siting, equipment type, and quality assurance criteria. NAMS monitors also must submit detailed quarterly and annual monitoring results to OAQPS. The drawback of the reporting mechanism is, mostly the sum or average of pollutant is informed to the sink. This lead to tremendous variation over original scenario and hence sudden unfavorable incidents of recent report of Delhi case - 16.10.2016 occurred [9,10].

4. SIMULATION STUDY

By considering the application scenario discussed in section 3 the simulation study using NS2 is carried out to collect air pollution data using various data aggregation functions. Table 2 shows the WSN setup parameters used in this study. Only the leaf nodes are sensors and have sensed values, all others are acting as intermediate nodes only to receive data from next level and transmit data to the previous level. We can get resultant aggregated data only from the sink node. Fig (2) shows initial node setup with assigned pollutant value. Fig (3) to (6) shows the node id with assigned value and sensed time with various aggregation functions applied namely total, average, maximum, and minimum. The advantage of these approaches is by applying various aggregation function it is possible to get aggregated value at sink node 0 at the particular time. The disadvantage is that if we preserve only the resultant aggregated value. It is not possible to know which node is sensed most and least pollutant value. Average and total functions are affected by extreme values may distort in these two functions and reduce its usefulness (Fig 3, 4).

Table 1 Data representation and in-network aggregation functions

TINA	Temporal coherency-aware in-network Aggregation
DADMA	Data Aggregation and Dilution by Modules Addressing
DAFC	Data Aggregation by means of Feedback Control
SDF	Synopsis Diffusion Framework
q-digest	Quantile Digest
DSC	Distributed Source Coding

Table 2 Parameters considered in this study

Network dimension	700 m x 700 m
Simulation duration	50s
No of maximum communication hop	Three
No of nodes	11 (0 to 10)
Nodes Location	Predefined Location
No of Leaf Nodes	6 (7,8,9,10,5,6)
Sink Node	1 (0)
Data sensed	Numeric data (May be Air,Water,Noise pollution level ,Temperature etc)
Network Model	Tree Structure

Accuracy and asymmetry of pollution data and quality of information about how much is an excess in quantities that are above standards is one of the major needs. In case of applying threshold values, it is possible to get the list of nodes those

having the values greater than threshold value. Fig (7) shows reporting the nodes that are having values greater than certain threshold value. But the values of all other nodes are unnoticed. So for further analysis and to find specific node to watch keenly are not possible in all five cases.

Tree Based Approaches	TAG, Directed Diffusion, PEGASIS, SDPA, DB-MAC,EADET
Cluster Based Approaches	LEACH,COUGAR, PEGASIS
Multipath Approaches	SynopsisDiffusion, DAG
Hybrid Data Aggregation	Tributaries and Deltas Approaches

Fig 1: Networking protocols and hierarchies for in-network aggregation

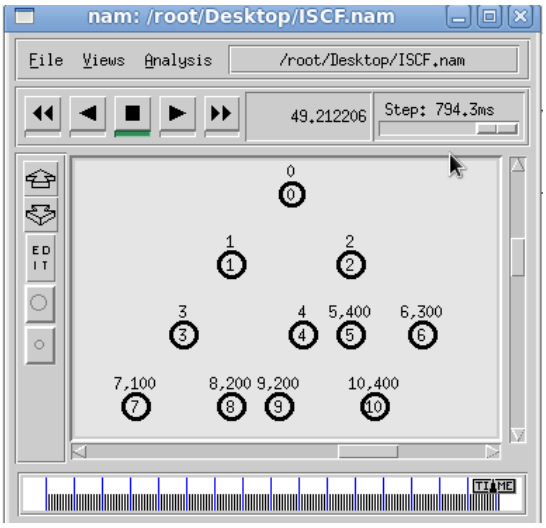

Fig 2: Initial node setup with assigned pollutant value

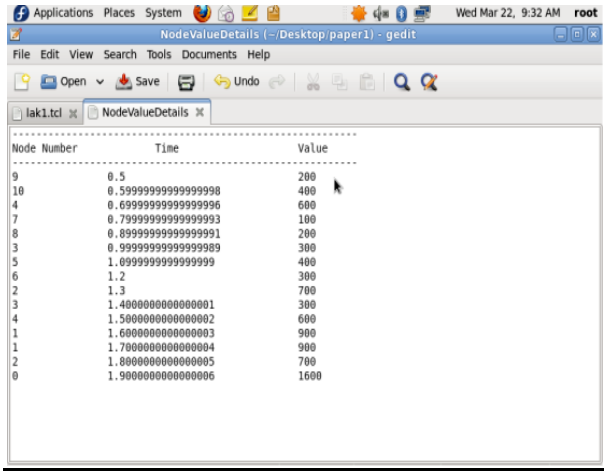

Fig 3: Data aggregation function: Total

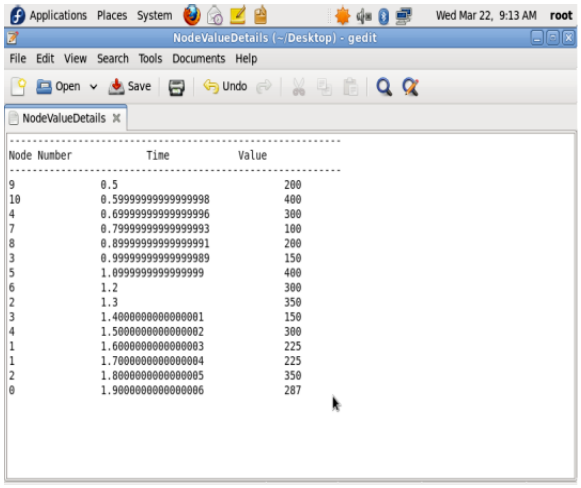

Fig 4: Data aggregation function: Average

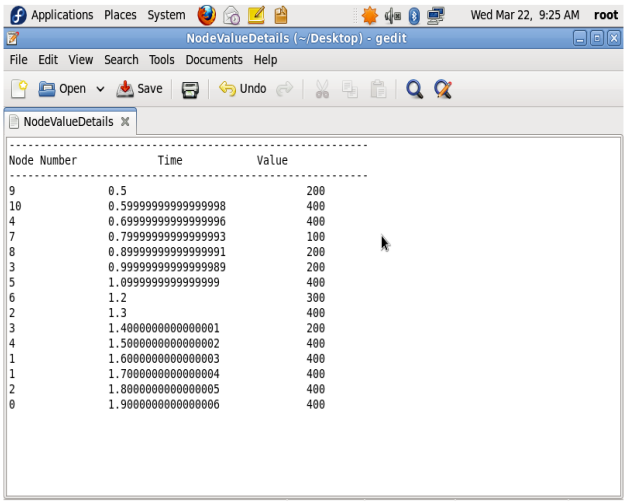

Fig 5: Data aggregation function : Maximum

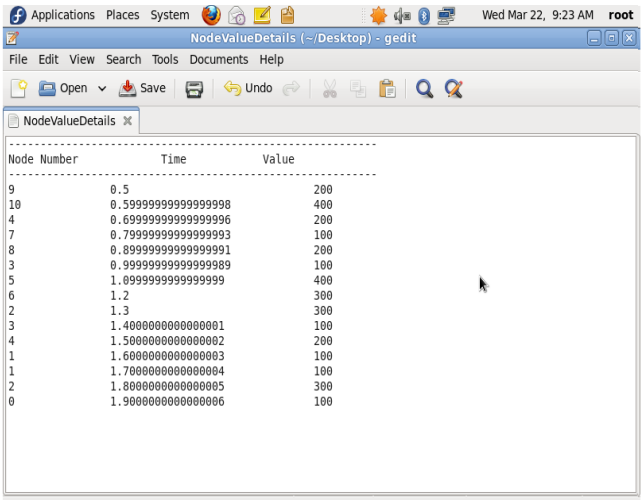

Fig 6: Data aggregation function: Minimum

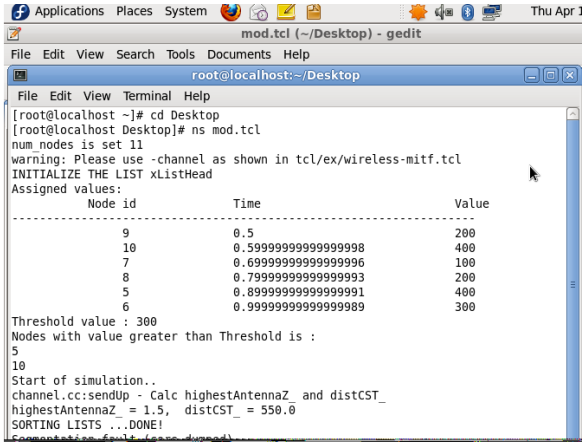

Fig 7: Reporting value greater than threshold value

5. EFFICIENCY OF THE DATA AGGREGATION FUNCTIONS

Data and information received from industries SPCB/CPCB have raised some insightful questions about the depth and quality of information available on pollution in the country. Even though data aggregation is to reduce power consumption, one important metric for the performance evaluation is Data Aggregation Quality. It is an interesting performance measures as it takes into account both the effectiveness in reducing the size of the data to be transmitted and the quality of the transmission. So, if we want to do effective analysis it is necessary to maintain all the sensed data, with node id, sensed time and sensed value by eliminating redundant value. To do better classification, it is tested that the sensed values are assigned randomly from 50 to 600 and nodes are classified based on sensed values as ambient nodes, tolerable nodes, and un tolerable nodes as shown in Fig (8).

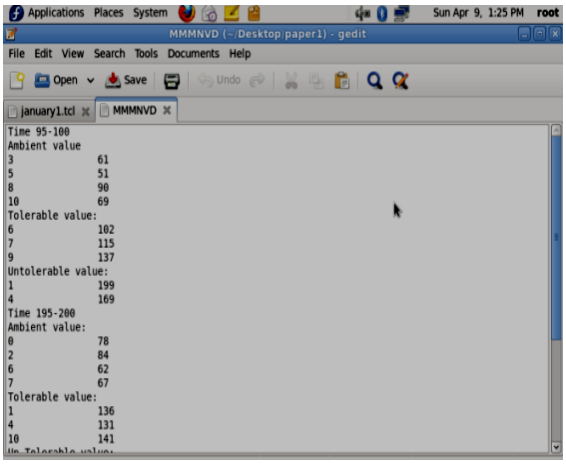

Fig 8: Three Classifications of nodes

A source node sends packet to the sink, in the frequency of period 100 ns. The format used in this study is source node id, sensed time and sensed value. The sink id is not included because there is only one sink, considered in this study.

The above format consists of two kinds of data, control information and user data (payload).The control information provides data the network needs to deliver the user data, for example source address, which identify the node uniquely in the network. The user data are time and the corresponding pollution reading collected by the source node.

6. CONCLUSION

The already available spot or sampling cannot give adequate data on the nature and magnitude of an air pollution problem. Factors related to continuous monitoring are number of communication, energy consumption and bandwidth. The possible alternative method of reporting is data aggregation. There is a lack of objective criteria for choosing appropriate aggregation method. Various methods of aggregation and their drawbacks and the efficient method needed for further analysis is discussed in this paper. Without doubt securing environment compliance has been a challenge and India's approach to pollution monitoring and prevention requires fresh thinking. This study is an attempt to secure better compliance by leveraging technology. Implementing new protocol design for energy efficient continuous pollution monitoring WSN is the next step to be considered.

7. ACKNOWLEDGEMENT

We gratefully acknowledge the helpful comments from the reviewers, which have improved the paper very significantly.

8. REFERENCES

[1] KAY ROMER AND FRIEDMANN MATTERN, ETH ZURICH "The design space of wireless sensor networks" IEEE Wireless Communications (Volume : 11, Issue : 6, Dec. 2004).

[2] Mohemed Younis [a,*], Kemal Akkay [b] , "Strategies and techniques for node placement in wireless sensor networks: A Survey", Elsevier Ad Hoc Networks 6 (2008) 621-655

[3] I.F.Akyildiz, W.Su,Y.Sankarasubramaniam, E.Cayirci,"A survey on sensor networks,IEEE Communication Magazine 40(80) (2002)104-112.

[4] S.Sasirekha [1*]and S.Swamynathan [2],"A comparative Study and Analysis of Data Aggregation Techniques in WSN", Indian Journal of Science and Technology,Vol 8(26), IPL0568, October 2015.

[5] E.Fasolo,M.Rossi,J.Widmer,and M.Zorzi," In-network aggregation techniqus for wireless sensor networks: A Survey",IEEE Wireless Communication., vol.14,no.2,pp.70-87,2007.

[6] Mousam Dagar [1] and Shilpa Mahajan [2],"Data Aggregation in Wireless Sensor Networks : A Survey",Interational Journal of Information and Communication Technology. ISSN 0974-2239 Volume 3,Number 3(2013),pp.167-174.

[7] Jyoti Rajput, Naveen Garg,"A Survey on Secure Data Aggregation in Wireless Sensor Network", IJARCSSE volume 4,Issue 5,May 2014.

[8] K.Akkaya and M.Younis, "A Survey of routing protocols in wireless sensor networks",Elsevier Ad Hoc Network Journal,vol.3,no.3,pp.325-349, May2005.

[9] TAMIL NADU POLLUTION CONTROL BOARD www.tnpcb.gov.in

[10] Central Pollution Control Board (CPCB) www.cpcb.nic.in