International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 2, February 2016

Data Filter Based Clustering method for Data Accuracy of Data Aggregation in WSN

[1] Sakthivel S [2] R. Nandhakumar
[1][2]Department of Computer Technology,
[1] [2] Anna University,
MIT, Chromepet, Chennai, India.
[1] sakthivale@gmail.com [2]nandhakumarr03@gmail.com

Abstract: **One data aggregation method in a wireless sensor network (WSN) is sending local representative data to the sink node based on the spatial clustering of sampled data. The recent spatial clustering models of sensor node's data are not appropriate due to its error rates, the representative data are inaccurate when compared with real data. Thus, an value based approach on node's data is followed to avoid larger deviations. A threshold value is set in each node by default, which excludes nodes with higher data differences from its cluster or aggregation. The excluded node is considered as an Isolated Sensor Node and sends its data directly to the sink, without participating in normal clusters. The proposed clustering method provides good data accuracy in data aggregation when compared to existing systems. The number of active nodes identified using the proposed method is lower than existing systems and it minimizes power consumption. This method provides more data accuracy when compared to existing clustering methods such as Pearson Correlation Coefficient clustering method, α-local spatial clustering method, and Data Density Correlation Degree Clustering method.**

Index Items : **WSN, aggregation, multihop, reprogramming.**

I. INTRODUCTION

Wireless Sensor Network (WSN) is a distributed in nature which consists of sensor nodes to sense some physical phenomenon or monitor the environmental conditions. Data transmission takes place in two ways. One is a single hop and another one is multihop to reach the base station through an intermediate node. Collected information from each sensor node transmits their data to the base station in a multihop manner, which increases the amount of redundant data during transmission. Due to their resource constraint, the rise in redundant information leads to more,

Nodes. To avoid this aggregation technique is used. Data aggregation takes place through either tree-based or cluster-based approach. On tree-based, the leaf node is sensor node and root is a base station. During the transmission of data from the leaf node to a base station, aggregation takes place at parent node. In cluster-based, sensor nodes in network and form cluster and a node selected as cluster head. Cluster head sends aggregate data and sends to a base station.

In WSN, Data aggregation is a process of collecting and aggregating the sensor node's data using aggregation approaches. The aggregated data is transfer to the sink node by selecting the efficient path. Effective utilization of energy in a node is a major problem in real time applications. A node in sensing and transmitting states consumes more energy, and its not possible to recharge all nodes. Even many nodes are deployed in remote areas Which can't be replaced in a easier manner. To reduce power consumption and to provide most appropriate data, clusters were formed. Cluster head collect the data from all the nodes in its cluster and sends aggregated value to the sink. In most cases data sent to cluster heads for aggregation doesn't has any filters, so there are chances for larger deviation, which increases error rates.

In figure 1, sensor nodes sense data and send it to their cluster head. Cluster head aggregate and send it to another cluster head. Another cluster head aggregates their node member's data and received previous cluster head data and similarly multichip to reach base station cause shrink in actually sensed data.

The summary of this section about aggregation and reprogramming problem in Wireless Sensor Network (WSN).

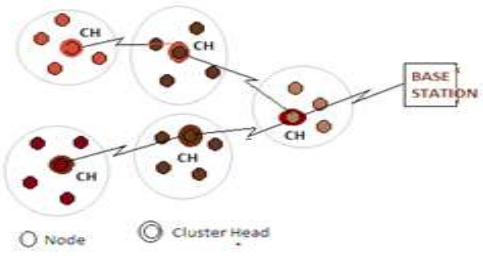

Fig. 1. Data Aggregation in WSN clusters.

II. RELATED WORK

Data Density Correlation Degree is a technique Highlights the problem in the recent spatial correlation models of sensor node's data which are not appropriate for measuring the correlation in a complex environment. In addition, the representative data are inaccurate when compared with real data. The data density correlation degree is a spatial correlation measurement that measures the correlation between a sensor node's data and its neighbor sensor node's data. Based on this correlation degree method was introduced for clustering, which is thus called a data density correlation degree (DDCD) clustering method[1]. The other famous correlation techniques are Pearson Correlation Coefficient and alpha α-local spatial clustering methods.

Performance Based Approach: The information contained in a transmission packet and makes the system prone to packet loss. Rather than retransmitting of lost packets, which leaves way for additional delay, a wireless system architecture has been developed which exploits the diversity of the wireless medium for robust and reliable operations. A class of aggregation functions were also discussed, where wireless broadcasting is an effective strategy to for reliability constraint. [2]

Aggregation Capacity: Random WSNs can be categorized into two types: random extended WSN and random dense WSN. The existing results about aggregation capacity were studied for dense WSNs, which includes random cases and arbitrary cases, under the protocol model (ProM) or physical model (PhyM). The first aggregation capacity scaling laws for random extended WSNs were proposed. They pointed out that unlike random dense WSNs, for random extended WSNs, the assumption made in ProM and PhyM that each successful transmission can sustain a constant rate is over-optimistic and unpractical due to transmit power limitation. [3]

Compressive data gathering (CDG) is a clustering method to provide energy efficient data aggregation in WSNs. A CDG requires efficient routing (forwarding) trees to gather encoded data from sensor nodes to the sink. Its also a decentralized method for the compressive data gathering problem (DCDG). This method allows each sensor node to locally make a decision in constructing and maintaining the forwarding trees and has minimal complexity and overhead with outstanding performance. [4]

Link Scheduling: To minimize power consumption, sleep scheduling algorithms can be used to turn the nodes to the sleep state whenever sensors are not in use and will wake up when requiered. In WSNs a Contiguous link scheduling problem is, when each node is assigned consecutive time slots so that the node can wake up only once in a scheduling period to fulfil its data collection task. It has been proved that the contiguous link scheduling problem in WSNs to be NP-complete. And it is considered as one of the best performing algorithm with theoretical performance bounds in both homogeneous and heterogeneous networks. [5]

DAT refers to Data gathering trees capable of performing aggregation operations. At present a large number of existing works focus on building DATs according to distinct user requirements using Deterministic Network Model (DNM). Constructing an Effective Load-Balanced Data Aggregation Tree (ELBDAT) under the PNM would be a solution. It also investigates three major problem, namely, the Load-Balanced Maximal Independent Set (LBMIS) problem, the Connected Maximal Independent Set (CMIS) problem, and the LBDAT construction problem. LBMIS and CMIS are well-known NP-hard problems and LBDAT is an NP-complete problem.[6]

Machine-to-Machine (M2M) paradigm provides a mechanism in which machines such as sensors, actuators, robots, and smart meter readers to communicate. M2M is an efficient and promising key enabling technology for the Cyber Physical Systems (CPS). The main objective is to develop a localized algorithm, where CPS nodes make decisions based on spatial and local knowledge.[7] Utilizing the clustering algorithm to form hierarchical network topology is the common method of implementing network management and data aggregation in wireless sensor networks. Many of the clustering algorithms heuristically determine the node energy, optimal node communication radius, and clustering radius, even though these parameters affect energy consumption of the entire network. In order to minimize energy consumption of the entire network, the optimal configuration by quantitative analysis when the nodes follow the random distribution was explained. [8]

A Swarm Intelligence based routing technique is widely used in network routing. LEACH (Low Energy Adaptive Clustering Hierarchy) is one of the most popular clustering algorithms. A novel routing approach based on ACO algorithm in Wireless Sensor Networks on which LEACH protocol is applied, to route the data packets in sensor networks to maximize energy efficiency and to increase the network lifetime. the redundant data sent by the sensors which are very much close to each other in the sensor network was reduced.[9]

In many network applications, the nature of traffic is of burst type. Often, the transient response of network to such traffics is the result of a series of interdependent events whose occurrence prediction is not a trivial task. A station-level stochastic time-domain method from which the network-level metrics are extracted was explained. This is in contrast to the previous efforts which have followed top down approaches i.e. through making top-view models of the whole network, they have tried to track the transient response of network to burst packet arrivals. [10]

The summary of this section is about previous spatial clustering and aggregation techniques in Wireless Sensor Network.

III. PROPOSED WORK

If all send its data to the sink, then more power would be consumed. So a cluster approach is formed and a cluster head is selected, which usually act as Representative Sensor Node. The nodes which did not participate in any clusters are called Isolated Sensor Node. The node which involves in transmission of one node's data to sink but doesn't participate in sensing that particular data it transmits is called a Member Sensor Node. Cluster Head collects all its participant node's data and aggregate that and send it to the sink via some MSNs. Error rate is a primary point of concern, when aggregating data for Sensor Nodes.

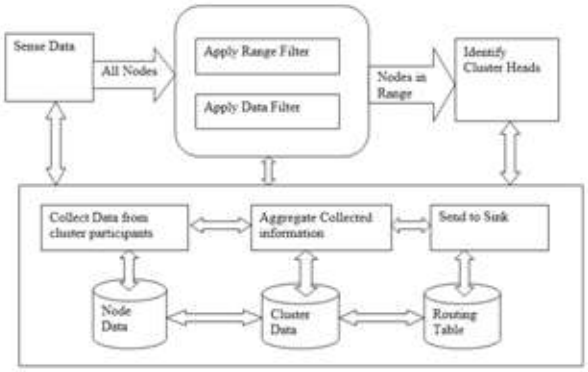

Fig. 2. System architecture

The proposed system is to reduce the error rates, in-turn increases the accuracy by adding a Data Filter on all nodes/Cluster Heads. A data threshold limit 'd' will be introduced as per out proposed model and any cluster nodes cannot consider the nodes with data when not in its data range+/- threshold.

A. SENSE DATA:

Sensing the required data is the most primary work of any sensor node. Sensed information is set to its nodes for filter references. Each node will have its own location specifically data and identification number to identify itself. In our implementation we will be using simulator to set data as sensed information.

B. FILTER DISTANCE:

We must consider the success replies and a threshold value will be identified, which will help in loss-less communication among neighboring nodes. We must take all nodes in its range and then filter only the nearest nodes for successful communication.

C. FILTER DATA:

With the result set provided by the above module, we should set data threshold using the below formula:

minAllowed = This node's data – d;
maxAllowed = This node's data + d;
where d is data threshold

D. IDENFITY CLUSTER HEAD:

It is necessary to count the number of nodes that each nodes contains. It will he helpful in identifying a cluster head. Count should be passed to other nodes in its range. The node which gets maximum nodes in its cluster will act as cluster head. The remaining nodes are considered as cluster members or Member Sensor Nodes. When racing condition arises, Two nodes having same number of participants, Node with lesser devia
tion will be considered as Cluster head.

E. NOTIFY PARTICIPANTS:

Here comes the decision part, all the nodes in the result set should be notified from the cluster head, so that it will send data to the cluster head while sensing.

F. AGGREGATE DATA:

At regular intervals sensor nodes are programmed to send sensed information to its head. When head receives the data from all the participants, it should take aggregate and send the data to sink. Nodes having no neighbors or no other participants in its cluster will be an Isolated Sensor Node and send its information to the Sink.

G. COUNT ACTIVE NODES:

To Check the energy utilization and performance it is also necessary to count the Number of active nodes, which sends it's sensed information to the sink. Depending on this performance can be evaluated. A new result set will be obtained which contains only valid notes after filtering process.

IV. RESULT AND DISCUSSION

Network simulator 2 is discrete event-driven network simulator. IEEE 802.15.4 provides MAC layer and Physical layer specification. The wireless model for static node, node with mobility. The simulation is set with thirty nodes deployed in the area of 700 m X 500 m and the routing protocol used is DSDV protocol, LL/ link layer. The channel used here is wireless channel. The language used for simulator is TCL (Tool Command Line) and C++. A node which covers more neighbor selected as the cluster head. Intra cluster head and inter cluster head aggregation are performed to reduce traffic need to be send among individual nodes to the base station. Accurate data obtained by increasing number of nodes. The Simulated Result shows the cluster formation in initial level and the initial clusters has two to three nodes on their cluster approximately which shows a clear filtering occurred on all the nodes. The initial

cluster head selection was 7 in numbers. It has 7 active nodes instead of 25. The reduction in more deviated participants shows the cluster will provide good accuracy than existing clustering algorithm for power optimization.

The simulated results shown the nodes with higher deviation of data value were kept to be isolated and accuracy of aggregation is better maintained.

Table.1. Simulation Parameters

PARAMETER	VALUE
Simulation area	700 m X 500 m
Channel	Wireless
Number of nodes	30
Packet size	512 bytes
Traffic type	TCP/UDP
Routing protocol	DSDV
Simulation time	300 sec

Table.2. Data Aggregation with Data Filter

Cluster Head ID	Aggregate Value	Node type
4	23	CH
12	27	CH
14	26	CH
19	22	CH
7	90	ISN
21	67	ISN
24	0	ISN

Table.3. Data Aggregation with out Data Filter

Cluster Head ID	Aggregate Value	Node type
4	35	CH
12	52	CH
14	26	CH
19	22	CH
24	0	ISN

V. CONCLUSION

Sensor networks are distributed in nature and resources constraints. Aggregation improve network lifetime by reduces amount of traffic need to be sent from an individual node to the base station. Sensor nodes senses data at frequent intervals of time and send to the cluster head. Cluster will be formed and select cluster head based on which node has more coverage neighbors. Communication takes place in one hop or multihop to reach the base station. Each sensor gathers the sensed information and passes them to aggregate in Cluster head. When followed this system, accuracy and optimal utilization can be achieved.

REFERENCES

[1]Fei Yuan, Yiju Zhan, and Yonghua Wang "Data Density Correlation Degree Clustering Method for Data Aggregation in WSN" IEEE SENSORS JOURNAL, VOL. 14, NO. 4, APRIL 2014 1089

[2]Dariush Ebrahimi and Chadi Assi "A Distributed Method for Compressive Data Gathering in Wireless Sensor Networks" IEEE COMMUNICATIONS LETTERS, VOL. 18, NO. 4, APRIL 2014

[3] Changhee Joo, Senior Member, IEEE, and Ness B. Shroff, Fellow, IEEE "On the Delay Performance of In-Network Aggregation in Lossy Wireless Sensor Networks" IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 22, NO. 2, APRIL 2014

[4] Cheng Wang, Changjun Jiang, Yunhao Liu, Senior Member, IEEE, Xiang-Yang Li, Senior Member, IEEE, and Shaojie Tang "Aggregation Capacity of Wireless Sensor Networks: Extended Network Case" IEEE TRANSACTIONS ON COMPUTERS, VOL. 63, NO. 6, JUNE 2014

[5] Junchao Ma, Student Member, IEEE, Wei Lou, Member, IEEE, and Xiang-Yang Li, Senior Member, IEEE "Contiguous Link Scheduling for Data Aggregation in Wireless Sensor Networks" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 7, JULY 2014

[6] Nandini. S. Patil, Prof. P. R. Patil, "Data Aggregation in Wireless Sensor Network". 2010 IEEE International Conference on Computational Intelligence and Computing Research.

[7] Ivan Stojmenovic, Fellow, IEEE, "Machine-to-Machine Communications With In-Network Data Aggregation, Processing, and Actuation for Large-Scale Cyber-Physical

Systems". IEEE INTERNET OF THINGS JOURNAL, VOL. 1, NO. 2, APRIL 2014

[8] M. Lee, and S. Lee, "Data Dissemination for Wireless Sensor Networks", in Proceedings of the 10 th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC'07).

[9] Tripti Sharma, Akanksha Dhawan, BijeshKumar, Rahul Singh Rathore, Karan Berry, Vishalakshi Gupta,."Ant Based Cluster Head Election Algorithm in Wireless Sensor Network to avoid redundancy". 2014 Fourth International Conference on Communication Systems and Network Technologies.

[10] H. Cam, S. Ozdemir, P. Nair, and D.Muthuavina shiappan, "ESPDA: Energy-Efficient and Secure Pattern-based Data Aggregation for Wireless Sensor Networks", in Proceedings of IEEE Sensor- The Second IEEE Conference on Sensors, Toronto, Canada, Oct. 22-24, 2003, pp. 732-736.

[11] Mohammad Sayad Haghighi, Member, IEEE, Yang Xiang, Senior Member, IEEE, Vijay Varadharajan, Senior Member, IEEE, and Barry Quinn, Member, IEEE,. "A Stochastic Time-domain Model for Burst Data Aggregation in IEEE 802.15.4 Wireless Sensor Networks". JOURNAL OF LATEX CLASS FILES, VOL. X, NO. Y, JUNE 2013

[12] Jing (Selena) He, Member, IEEE, Shouling Ji, Student Member, IEEE, Yi Pan, Senior Member, IEEE, and Yingshu Li, Senior Member, IEEE, "Constructing Load-Balanced Data Aggregation Trees in Probabilistic Wireless Sensor Networks" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 7, JULY 2014.

[13] S. Lindsey and C. Raghavendra, "PEGASIS: Powerefficient gathering in sensor information systems," in 2010 IEEE International Conference on Computational Intelligence and Computing Research Proceedings of IEEE AerospaceConference, vol. 3, Mar. 2002, pp. 1125–1130.

[14] Ying Liao, Jianjing Shen, Member, IEEE, Yi Lin, and Changlin Zhou, Senior Member, IEEE,"Quantitative Analysis of Network Configuration in Randomized Distribution Wireless Sensor Networks". IEEE SENSORS JOURNAL, VOL. 14, NO. 6, JUNE 2014

[15] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed Diffusion for Wireless Sensor Networking", IEEE/ACM Transactions on Networking, Vol. 11, no. 1, Feb 2003.

[16] Mohammad Sayad Haghighi, Member, IEEE, Yang Xiang, Senior Member, IEEE, Vijay Varadharajan, Senior Member, IEEE, and Barry Quinn, Member, IEEE,. "A Stochastic Time-domain Model for Burst Data Aggregation in IEEE 802.15.4 Wireless Sensor Networks". JOURNAL OF LATEX CLASS FILES, VOL. X, NO. Y, JUNE 2013

[17] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks", (MobiCom 2000) pp 56-67.

[18] K. Dasgupta, K. Kalpakis, and P. Namjoshi, "An Efficient Clustering-based Heuristic for Data Gathering and Aggregation in Sensor Networks", IEEE 2003.

[19] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-Network Aggregation Techniques for Wireless Sensor Networks: A Survey", IEEE Wireless communication 2007.

[20] M. Lee and V.W.S. Wong, "An Energy-aware Spanning Tree Algorithm for Data Aggregation in Wireless Sensor Networks," IEEE PacRrim 2005, Victoria, BC, Canada, Aug. 2005.

[21] M. Ding, X. Cheng, and G. Xue, "Aggregation tree construction in sensor networks," in Proc. of IEEE VTC'03, Vol. 4, Orlando, FL, Oct. 2003.

[22] "The Design Space of Wireless Sensor Networks" by Kay R¨omer and Friedemann Mattern http://www.vs.inf.ethz.ch/publ/papers/wsndesignspace.pdf

[23] http://en.wikipedia.org/wiki/Wireless_Sensor_Network