International Journal of Engineering and Techniques - Volume 1 Issue 4, July-Aug 2015
RESEARCH ARTICLE OPEN ACCESS

Survey: Robust Data Aggregation Techniques in Wireless Sensor Networks

Doddappa Kandakur
M.Tech Student, Dept. of CSE
SIT,Tumkur,

Ashwini B P
Asst. Professor, Dept. of CSE
SIT,Tumkur

Abstract:

Data aggregation in important issue in WSN's. Because with the help of data aggregation; we are reduce energy consumption in the network. In the Ad-hoc sensor network have the most challenging task is to maintain a life time of the node. due to efficient data aggregation increase the life of the network. In this paper, we are going to provide the information about the type of the network and which data aggregation algorithm is best. In big scale sensor network, energy economical, data collection and query distribution in most important.

Keywords — data aggregation; wireless sensor network

I. INTRODUCTION

Wireless sensor networks comprise of sensor hubs with detecting and correspondence capabilities [. The greater part of the vitality utilization is because of data transmission. For that we have a tendency to apply Data aggregation approach on the detected data by the sent sensor hubs.

A large portion of the specialists concentrate on data aggregation issues in vitality compelled sensor networks. The principle objective of data aggregation algorithms is to accumulate and aggregate data in a vitality productive way so that system lifetime is improved. I hypothetically investigate various algorithms on the premise of execution measures, for instance, lifetime, dormancy and data accuracy.

DATA AGGREGATION: AN OVERVIEW

The data aggregation is a technique used to solve the implosion and overlap problems in data centric routing. Data coming from multiple sensor nodes is aggregated as if they are about the same attribute of the phenomenon when they reach the same routing node on the way back to the sink. Data aggregation is a widely used technique in wireless sensor networks. The security issues like data confidentiality and integrity, in data aggregation become vital when the sensor network is deployed in a hostile environment. Data aggregation is a process of aggregating the sensor data using aggregation approaches.

THE NEED FOR DATA AGGREGATION

Sensor nodes are conveyed in remote situations to a multi-hop WSN more than an extensive variety of range. Once in a while do the clients have worldwide data on the sensor nodes' distribution. That is the reason when clients solicitation state-Based sensor readings of the attributes like temperature and humidity in a discretionary territory, systems may endure the flighty overwhelming activity. This issue needs data aggregation to conform to client prerequisites and Manage covered aggregation trees of multiple clients efficiently. Numerous down to earth applications like ecological checking, military applications, logical examination and so forth. are investigating the utilization of WSNs. Such applications oblige exchanging a tremendous measure of importance, sensed data starting with one purpose of the system then onto the next. Since WSNs are for the most part furnished with low power batteries, battery life is a noteworthy requirement in any constant application. This requires the utilization of energy efficient data dissemination protocols for aggregation of the sensed data. Nodes of a WSN in close closeness typically hold comparable data because of a property called spatial correlation

In a perfect data aggregation scheme, every sensor ought to be spending the same measure of energy in every data gathering round. A data aggregation scheme is energy productive in the event that it boosts the usefulness of the network. Here accepted all sensors are just as essential, we ought to minimize the energy consumption of every sensor. When an inquiry is sent by the BS to a sensor, the first step took after is to handle the question. This is trailed by data collection from sources and aggregation of that data.

Data aggregation obliges an alternate forwarding standard contrasted with classic routing. Classic routing Protocols commonly forward data along the shortest path to the destination (regarding some predefined metric). In the event that, nonetheless, Most of the researcher is interested in aggregating data to minimize energy use, hubs ought to course packets taking into account the parcel content and pick the following hop to advance in-network aggregation. This sort of data forwarding is frequently alluded to as data centric routing.

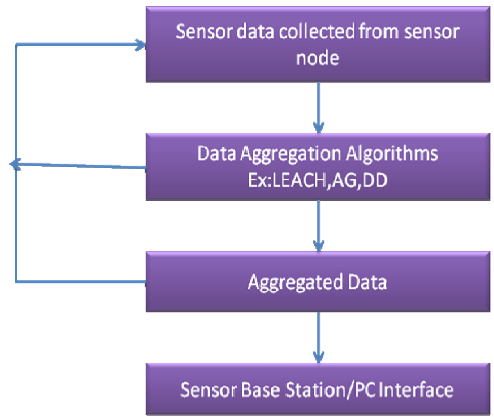
The execution measures of data aggregation algorithms such Network Lifetime, Latency, and Data Accuracy are portrayed beneath.

System lifetime: The system lifetime is characterizing the quantity of data fusion rounds. Till the predefined Percentage of the aggregate nodes bites the dust and the rate rely on upon the application.

Inertness: Latency is characterized as the delay included in data transmission, routing, and data aggregation. It to be able measured as the time delay between the data packet got at the sink and data created at the source node.

Data precision: It is assessing of proportion of aggregate number of perusing got at the base station (sink) to the aggregate number of produced.

General architecture of the data aggregation algorithm



General architecture of the data aggregation algorithm shows how data flow from sensor (sensed data) to Sink

II. DATA AGGREGATION BASED NETWORKS

Here we separation sorts of aggregation based network as takes after

1. Level networks
2. Various levelled networks

.

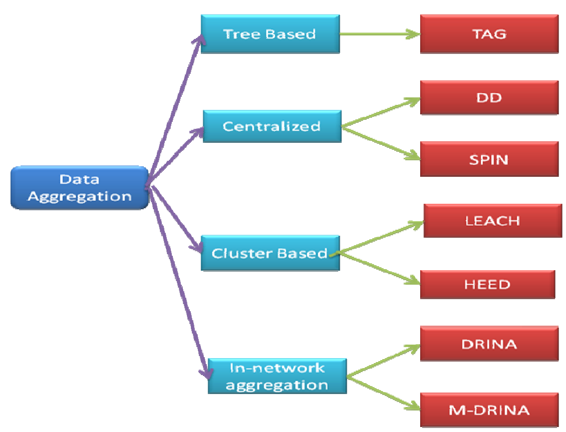


Figure2. Hierarchical Structure for data aggregation

A. FLAT NETWORKS

Level networks assumes critical part in wireless sensor network, in which every sensor hub has an equivalent battery control and plays the same kind of part in a network. In such sort of networks, data aggregation must be done in data centric routing way, where the sink for the most part sends a data parcel to the sensor hubs, for example, flooding. In the flooding sensors which have data coordinating the data bundle and transmit reaction data parcel back to the sink. The decision of a specific correspondence convention relies on upon the particular application within reach.

Most useable techniques are

In future we utilized these techniques some essential qualities.

B. HIERARCHICAL NETWORKS

The whole correspondence and reckoning weight at the sink in level network, that is the reason parcel of energy is expended. Henceforth, in perspective of adaptability and energy efficiency, a few progressive data aggregation methodologies have been proposed. Various levelled data aggregation includes data combination at extraordinary hubs, which diminishes the quantity of messages transmitted to the sink. This enhances the energy efficiency of the network.

As attributes of calculations concern on energy utilization of perspective, followings are oftentimes utilizes data aggregation as a part of wireless sensor network.

1. Low Energy Adaptive Clustering Hierarchy (LEACH)[2].
2. Mixture Energy Efficient Distributed Clustering Approach (HEED)
3. Grouped diffusion with dynamic data aggregation (CLUDDA).

III. ARCHITECTURES OF DATA AGGREGATION

Centralized Approach: In this approach only one sensor node play a role of aggregator node and all other sensor nodes are connected to that aggregator node. All other sensor nodes sense the data and transmit to the aggregator node which is called centralized node.

There are huge loads on that aggregator node, so there is a need for more energy and security on that aggregator node because all data is on the centralized aggregator node

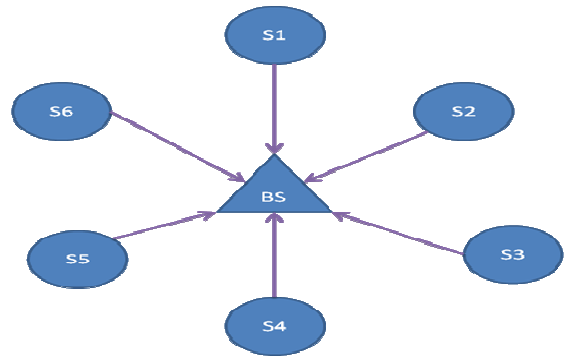


Figure3. Centralized approach for data aggregation in WSN

CENTRALIZED DATA AGGREGATION

Data is gathering at centre node in centralized data aggregation technique. For this process it takes the help of shortest path using a multi-hop wireless protocol. The sensor nodes send the data packets to a centre node, which is the powerful node. The leader aggregates the data which can be queried. Each intermediate node has to send the data packets addressed to leader from the child nodes. So a large number of messages have to be transmitted for a query in the best case equal to the sum of external path lengths for each node. Ex. DD, SPIN.

1. DD (Direct Diffusion)

It is data-centric protocol which sense data with the help of attribute-value pairs such as duration, geographical area, and interval.

2. SPIN (Sensor Protocol for Information via Negotiation)

It uses meta-data or high level descriptors. Meta-data are exchanged among sensors via a data advertisement mechanism before transmission.

Decentralized Approach: In this approach all sensor nodes performs aggregator function to the sensed data .In this approach there is no single centralized aggregator node but all nodes have same priority to aggregate the sensed data. In this approach all sensor nodes are connected to their neighbour node. This methodology has the benefit of more scalability, dynamic change node failure in the wireless sensor network.

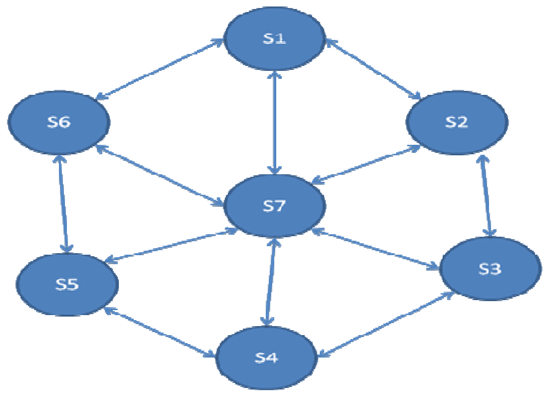


Figure4. Decentralize approach for data aggregation in WSN

Tree-Based Approach: The tree based approach is defining aggregation from constructing an aggregation tree. The form of tree is minimum spanning tree, sink node consider as a root and source node consider as a leaves. Information flowing of data start from leaves node up to root means sink(base station).Disadvantage of this approach, as we know like wireless sensor network are not free from failure .in case of data packet loss at any level of tree, the data will be lost not only for single level but for whole related sub tree as well. This approach is suitable for designing optimal aggregation techniques. TAG approach is better compare to existed approaches so we consider this, but TAG having some advantages and limitation we can find Table 1.0

TAG: The Tiny Aggregation (TAG) approach is a data-centric protocol. It is based on aggregation trees and specifically designed for monitoring applications. This means that all nodes should produce relevant information periodically. Therefore, it is possible to classify TAG as a periodic per hop adjusted aggregation approach. The implementation of the core TAG algorithm consists of two main phases:

- The distribution phase, where queries are disseminated to the sensors
- The collection phase, where the aggregated sensor readings are routed up the aggregation tree for the distribution phase, TAG uses a tree based routing scheme rooted at the sink node. The sink broadcasts a message asking nodes to organize into a routing tree and then sends its queries. In each message there is a field specifying the level, or distance from the root, of the sending node (the level of the root is equal to zero). Whenever a node receives a message and it does not yet belong to any level, it sets its own level to be the level of the message plus one. It also elects the node from which it receives the message as its parent. The parent is the node that is used to route messages toward the sink.

Cluster-Based Approach: In energy-constrained sensor networks of large size, it is inefficient for sensors to transmit the data directly to the sink In such scenarios, Cluster based approach is hierarchical approach. In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster-heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to base station (sink). Recently, several cluster-based network organization and data-aggregation protocols have been proposed for the wireless sensor network. Figure shows a cluster-based sensor network organization.

CLUSTER BASED DATA AGGREGATION

This methodology likewise comprises of hierarchical organization of nodes where nodes are separated into clusters with some unique nodes to regard as a cluster head are chosen to total data and advances it to the sink node.

Ex. LEACH, HEED

LEACH: Low-Energy Adaptive Clustering Hierarchy (LEACH) is a self-sorting out and versatile clustering protocol utilizing randomization to equitably convey the vitality consumption among the sensors. Clustered structures are misused to perform data aggregation where cluster heads go about as aggregation focuses The protocol works in rounds and defines two main phases:

- A setup stage to sort out the clusters
- A steady-state phase that deals with the actual data transfers to the sink node.

IN-NETWORK AGGREGATION

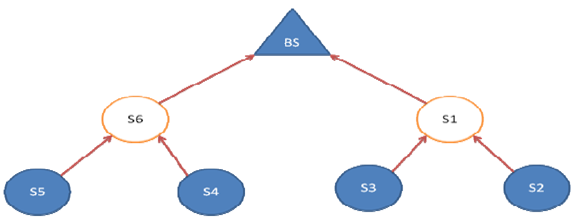
We can distinguish between two approaches:

• 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. As an example, assume that a node receives two packets from two different sources containing the locally measured temperatures. Instead of forwarding the two packets, the sensor may compute the average of the two readings and send it in a single packet.

• In-network aggregation without size reduction refers to the process of merging packets coming from different sources into the same packet without data processing: assume receiving two packets carrying different physical quantities (e.g., temperature and humidity). These two values cannot be processed together, but they can still be transmitted in a single packet, thereby reducing overhead.

The first approach is better able to reduce the amount of data to be sent over the network, but it may also reduce the accuracy with which the gathered information can be recovered at the sink. After the aggregation operation, it is usually not possible to perfectly reconstruct all of the original data. This actually depends on the type of aggregation function in use (i.e., lossy or lossless.)

The second approach instead preserves the original information (i.e., at the sink, the original data can be perfectly reconstructed). Which solution to use depends on many factors including the type of application, data rate, network characteristics, and so on. Both of the above strategies may involve the treatment of data at different network layers.



The idea of the INA is to aggregate the data required for the determination of the derivatives as close to the source as possible, instead of transmitting all sensed values through the entire network.

Ex. DRINA, M-DRINA

i) DRINA (Data Routing In-Network Aggregation)

DRINA algorithm is cluster-based approach. It works in three phases. In Phase 1, the hop tree is built when sensor nodes communicate with the sink node and the sink node starts building the hop tree that get used by coordinators for data forwarding purposes. In phase 2 cluster formation and cluster-head election is done among the nodes that detect the occurrence of a new event in the network. Finally, Phase 3 is responsible for both setting up a new route for the reliable delivering of packets and updating the hop tree.

SYNOPSIS DIFFUSION IN NETWORKING

This is a novel In-network aggregation framework that empowers robust, very exact estimations of duplicate-sensitive aggregates. The basic approach is to use best effort, multi-path routing together with duplicate-insensitive in- network aggregation schemes.

Synopsis diffusion performs in-network aggregation. The partial result at a node is represented as a synopsis a small digest (e.g., histogram, bit-vectors, sample, etc.) of the data. The aggregate computation is defined by three functions on the synopses are Synopsis Generation, Synopsis Fusion and Synopsis Fusion.

In spite of the fact that the synopsis diffusion framework is autonomous of the fundamental topology, to make it more concrete, overlay Ring topology, called Rings, which arranges the nodes into an arrangement of rings around the querying node.

This approach is robust in data aggregation (like count, sum and average) with high data security and overcome the node failure problem.

PEFORMANCE MEASURES OF DATA AGGREGATION

Below table show comparison of different data aggregation functions (algorithms):

Algorithm/ Framework	Type	Advantages	Disadvantages
TAG	Tree based	Ability to tolerate disconnections and loss.	Network life time is limited, Not covers node failure.
DD	Centralized	It extends the network lifetime	Usability less due to discontinuity in data delivery
LEACH	Cluster based	Less energy Consumption, Network lifetime increases	Limited uses because not suitable for large network, Limitation of the Node's communicate on range
DRINA	In-Network	Data Integrity, Less energy Consumption	Dynamically not rotating cluster Head
Synopsis diffusion	In-Network	Data high security, Less communication overhead,	But compromise with high priority attackers in network.

		Less energy Consumption	Not Covers DOS attacks

CONCLUSION

This method is one of the dynamic probabilistic schemes using the numbers of child and sibling nodes. The node has higher retransmission probability if it has the more child nodes and the less sibling nodes. This method needs initial overhead due to the first five instances of flooding, but it will outperform all the other methods in the aspect of the number of retransmissions as we discussed in the working example.

ACKNOWLEDGMENT

I would like to thanks Prof. Ashwini B P, Dept. of CSE, Siddaganga Institute of Technology Tumkur for guiding and providing valuable suggestions to complete this paper.

REFERENCE

[1] C. Shen, C. Srisathapornphat, and C. Jaikaeo, "Sensor information networking architecture and applications," *IEEE*

[2] E.J. Duarte – Melo, and M. Liu, "Data-gathering wireless sensor networks: organization and capacity," Computer networks: The *International Journal of Computer and Telecommunications Networking*, vol. 43, no.

[3] Fabbri, F, Buratti, C. ; and Verdone, R." A multi-sink multi-hop wireless sensor network over a square region: Connectivity and energy Consumption issues" GLOBECOM Workshops, 2008 IEEE, pp1 - 6

[4]M. Kuorilehto, M. Hännikäinen, and T. D. Hämälä inen, "A survey of application distribution in wireless sensor networks," EURASIP Journal on Wireless Communications and Networking, Vol. 2005, October 2005.

[5] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, and D. Moore,"Environmental wireless sensor networks," *Proc. IEEE*, vol. 98, no. 11, pp. 1903–1917, Nov. 2010.

[5]Ankit Tripathi, Sanjeev Gupta, and Bharti 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

[6]Prakashgoud Patil, and Umakant Kulkarni "Analysis of Data Aggregation Techniques in Wireless Sensor Networks" IJCEM International Journal of Computational Engineering & Management, Vol. 16 Issue 1, January 2013,ISSN (Online): 2230-7893

[7] K. Akkaya and M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks," *Elsevier Ad Hoc Network J.*, vol. 3, no. 3, May 2005, pp. 325–49.

[8] K. Akkaya and M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks," *Elsevier Ad Hoc Network J.*, vol. 3, no. 3, May 2005, pp. 325–49.

[9] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data Gathering Algorithms in Sensor Networks using Energy Metrics," *IEEE Trans. Parallel Distrib. Sys.*, vol.13, no. 9, Sept. 2002, pp. 924–35.

[10] Y. Yao and J. Gehrke, "Query Processing for Sensor Networks," *ACM CIDR 2003*, Asilomar, CA, Jan. 2003.

[11] S. Madden *et al.*, "TAG: a Tiny AGgregation Service for Ad Hoc Sensor Networks," *OSDI 2002*, Boston, MA, Dec. 2002.

[12] Y. Xu, J. Heidemann, and D. Estrin, "Geographic-Informed Energy Conservation for Ad Hoc Routing," *ACM/SIGMOBILE MobiCom 2001*, Rome, Italy, July 2001.

[13] G. Di Bacco, T. Melodia, and F. Cuomo, "A MAC Protocol for Delay-Bounded Applications in Wireless Sensor Networks," *Med-Hoc-Net 2004*, Bodrum, Turkey, June 2004.

[14] S. Nath *et al.*, "Synopsis Diffusion for Robust Aggregation in Sensor Networks," *ACM SenSys 2004*, Baltimore,MD, Nov. 2004.

[15] A. Manjhi, S. Nath, and P. B. Gibbons, "Tributaries and Deltas: Efficient and Robust Aggregation in Sensor Network Stream," *ACM SIGMOD 2005*, Baltimore, MD,June 2005.

[16] I. Solis and K. Obraczka, "The Impact of Timing in Data Aggregation for Sensor Networks," *IEEE ICC 2004*, Paris, France, June 2004.

[17] F. Hu, C. Xiaojun, and C. May, "Optimized Scheduling for Data Aggregation in Wireless

Sensor Networks," *IEEE ITCC '05*, Las Vegas, NV, Apr. 2005.

[18] X. Jianbo, Z. Siliang, and Q. Fengjiao, "A New In-network Data Aggregation Technology of Wireless Sensor Networks," *IEEE SKG '06*, Guilin, China, Nov. 2006.

[19] E. Cohen and H. Kaplan, "Spatially-Decaying Aggregation Over a Network: Model and Algorithms," *ACM SIGMOD '04*, Paris, France, June 2004.

[20] A. Sharaf *et al.*, "Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks," *VLDBJ.*, vol. 13, no. 4, Dec. 2004, pp. 384–403.

[21] T. He *et al.*, "AIDA: Adaptive Application-Independent Data Aggregation in Wireless Sensor Networks," *ACM Trans. Embedded Computing Systems*, vol. 3, no. 2, May 2004, pp. 426–57.

[22] E. Cayirci, "Data Aggregation and Dilution by Modulus Addressing in Wireless Sensor Networks," *IEEE Commun. Lett.*, vol. 7, no. 8, Aug. 2003, pp. 355–57.

[23] D. Petrovic *et al.*, "Data Funneling: Routing with Aggregation and Compression for Wireless Sensor Networks," *IEEE SNPA '03*, Anchorage, AK, May 2003.

[24] M. Riedewald, D. P. Agrawal, and A. El Abbadi,"pCube: Updateefficient Online Aggregation with Progressive Feedback and Error Bounds," *IEEE SSDBM 2000*, Berlin, Germany, July 2000.

[25] L. Huang *et al.*, "Probabilistic Data Aggregation in Distributed Networks," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2006-11, Feb.6, 2006; http://www.eecs.berkeley.edu/Pubs/TechRpts/ 2006/EECS-2006-11.html

[26] X. Wu and Z. Tian, "Optimized Data Fusion in Bandwidth and Energy Constrained Sensor Networks," *IEEE ICASSP '06*, Toulouse, France, May 2006.

[27] N. Shrivastava *et al.*, "Medians and Beyond: New Aggregation Techniques for Sensor Networks," *ACM SenSys '04*, Baltimore, MD, Nov. 2004.

[28] A. Bezenchek, M. Rafanelli, and L. Tininini, "A Data Structure for Representing Aggregate Data," *IEEE SSDBM '96*, Stockholm, Sweden, June 1996.