DATA AGGREGATION ALGORITHMS IN WIRELESS SENSOR NETWORKS: A REVIEW

MR. PABBOJU RAMESH ,
Professor, Department of C.S.E
Mahaveer Institute of Science and Technology, Bandlaguda, Hyderabad, T.S,
pramesh.448@gmail.com

DR. R.NAKKEERAN, HOD,
Department of C.S.E
Mahaveer Institute of Science and Technology, Bandlaguda, Hyderabad, T.S,
sudhandhiram64@gmail.com

Abstract: *In the wireless sensor network (WSN) the multi hop designed based on the application requirement of the node in sensor. Main aim of wireless sensor network is to provide a collected data to sink and monitors the given area of interest during the physical phenomenon using sensors. If the nodes are deployed in a hostile environment the overcoming of energy constraints by recharging and replacing the batteries in the nodes of WSN becomes impossible and thereby to keep the network alive for a long time as maximum the communication has to be performed with the load balancing in wireless sensor networks . The WSN is built by the communication, energy and computational constraint nodes. To protect from disaster the forest fire detection and sink which monitors the demand reception of data with the bounded delay are used which is a time critical applications. Thus requires a protocol to be designed for providing the data to sink with bounded delay and enhance the lifetime of network, data accuracy. In this paper, we present a survey of data aggregation algorithms in wireless sensor networks and possible future research directions.*

Keywords: Wireless Sensor Networks, data aggregation, lifetime, latency and data accuracy, Modified Cuckoo search (MCS) Algorithm, Modified time on task (MTOT) method, Gravitational search algorithm (GSA)

INTRODUCTION

Wireless sensor networks (WSN) mostly uses aggregated node to collect data from sensor nodes in the network. The data aggregation process creates major problem in normal data transmission such as, energy consumption, and delay. Many researchers have handled many time data aggregation problem and some aggregation functions used to aggregate multiple data into one or every data into one, which reduce the energy consumption. Data aggregation is used for grouping the data from multiple sensor nodes by avoiding the redundant data transmission and that data has been sent to the base station in single route. Aggregation process reduces number of redundant transmission and thus it improves the network performance and also provides bandwidth utilization. In WSN, traditionally nodes send data individually when the base station demands for network. Instead of that there is a special node called aggregator which used to collects statistics information from its neighbouring stations, adds them and forward that combined information to the base station in multi-hop manner. There are two different nodes involved in the wireless sensor network such as the normal sensor node and sink node also referred as aggregator node. At first, we have to compute the data aggregation node in the sensor networks based on the lifetime. The lifetime of sensor nodes depends on time varying factors, such as consumed energy for sending data from the leaves to aggregated node, queuing delay during aggregation, and tree depth delay. In the sensor network there are mostly repeated and interrelated data produced from the neighbouring sensor nodes. Here used a technique named data aggregation. The sensor node uses substantial energy for transmitting a huge data. But also the high amount of data transmission leads to data congestion and data collision. Thus the technique is used for removing a redundant data. Data aggregation is a technique which aggregates a data from the multiple sensor nodes without data redundancy and sink node receives a complex data. There are three methods in data aggregation. First method is clustering aggregation which collects a data and aggregated then send to a sink node [7] [8]. Secondly the data get aggregated with the intermediate node named hop by hop aggregation [9]. The next is a partial aggregation which the threshold of energy and time satisfies the data aggregation [10].Anyhow all these three methods contain some drawbacks in processing. In the first method cluster head always consumes more energy compared to others. The cluster head selecting algorithm are used for choosing the new cluster head thus more energy and time get wasted [8]. In the hop by hop aggregation the drawbacks are causing long delay during transmission of data and packets due to unbalanced energy consumption [11]. While comparing the partial aggregation is better with other two methods because it uses the time scheduling technique. In this node has a aggregating function which time is allotted to collect a data. In this the energy consumption is not always considered for transmitting nodes and the time period is decided by the reliability of requested data and a given threshold time in the network [12-16]. In the network the lifetime of network is based on the nodes lifetime but when the node losses its energy the reorganization is performed by the network node. So that however this aggregation technique reduces a delay but the lifetime of network cannot be increased and has unbalanced nodes.

I. DATA AGGREGATION BASED NETWORKS

1. Flat networks
2. Hierarchical networks

A. Flat network based Aggregation

A tree rooted approach is commonly used for the efficient data aggregation at the sink. It uses a data centric protocol for aggregation of data from leaf nodes. The data gathering tree is constructed in order to enhance the network life, where nodes used have heterogeneous and adjustable power level. It helps to find the upper bound on network life time by rearranging the tree structure based on heavily loaded nodes. This approach is suitable for network with number of nodes less than 25. Energy efficient spanning tree algorithm based on the residual energy is proposed. It has the advantage of increase in average node life time, since root to sink is selected based on the highest availed residual energy of node . ACO (Ant colony algorithm) represents the opportunistic aggregation for formation of low latency paths. Energy wise opportunistic aggregation near the source is not optimal. The flat networks in WSN uses tree structure, where all the leaf nodes forward the data to its parents and then it is rooted towards the sink. Tree approach causes much transmission delay and reduced packet delivery ratio since any one node failure in the root blocks the data.

B. HIERARCHICAL NETWORKS

All the communication and computation burden at the sink in flat network, that's why lot of energy is consumed. In the hierarchical network, In which data aggregation data has to be done at special nodes, with the help of these special node we can reduce the number of number of data packet transmitted to the sink. So with this network improves the energy efficiency of the whole network.

Table 1 Hierarchical Network vs. Flat Networks

Hierarchical network	Flat Network
Data aggregation performed by cluster heads or leader node	Data aggregation is performed by different nodes along the multi-hop path
Overhead involved in cluster or chain formation throughout the network	Data aggregation routes are formed only in regions that have data for transmission
Even if one cluster head fails, the network may still be operation	The failure of sink node may result in the breakdown of entire network
Lower latency in involved since sensor nodes perform short rang transmission to the cluster head	Higher latency is involved in the data transmission to the sink via multihop path.
Routing structure is simple but not necessarily optimal	Optimal routing can guaranteed with additional overhead
Node heterogeneity can exploited by assigning high energy nodes as cluster heads	Does not utilize node heterogeneity for improving energy efficiency.

Table 2 protocols for Hierarchical and Flat networks

Protocol Name	Flat Network	Hierarchical Network
LEACH		▪
PEGASIS		▪
HIERARCHICAL PEGASIS		▪
SPIN	▪	
DIRECTED DIFFUSION	▪	
TEEN		▪
APTEEN		▪
ROUMAR ROUTING	▪	
GRADIENT BASED ROUTING	▪	
ENERGY AWARE ROUTING FOR CLUSTER BASED SENSOR NETWORK		▪
CADR(CONSTRAINED ANISOTROPIC DIFFUSION ROUTING)	▪	
ACQIRE(ACTIVE QUERY FORWARDING IN SENSOR NETWORK)	▪	
ENERGY AWARE ROUTING	▪	

B.1 Cluster based Aggregation

To overcome the problem of transmission delay and loss of data due to node failures in the root to sink, cluster based aggregation is used. The CH aggregates the information from nodes and performs the additive or divisible aggregation function. Different cluster based protocols are explored in. In CWCG (Cluster wide correlated grouping) hybrid structure for data aggregation is proposed which uses the concept of temporal and spatial grouping of nodes. It provides reduced transmission cost but increases the latency. In ADA (Adaptive data aggregation) temporal aggregation degree is controlled by the reporting frequency of the events at the sensor nodes and spatial degree of aggregation is set by the aggregation ratio at CH. Sink has the central control for aggregating the data by sending temporal and special degree. It provides improvement in the desired reliability than observed. In Hybrid approach is presented for static and dynamic clustering, protocol extends towards the adaptive nature based on the velocity of target. Dynamic clustering shows better performance when velocity of target is high but degrade with imperfect data aggregation. Select cast considers the optimal tradeoff between aggregation throughput and gathering efficiency. It uses the spatial correlation between data collected from the sensor with lower bound as threshold value for the gathering efficiency and throughput. The data aggregation uses perfectly compressible function such as mean and max. CTEPEDCA (Cluster-based and Tree based Power efficient Data collection And Aggregation protocol), is based on the minimum spanning tree (MST) routing, where only one CHs communicates with base station. It shows improvement in network life time with small time complexity, if numbers of clusters are increased. CPDA (Cluster-based Private data Aggregation) bridges the gap between collaborative data collection and data privacy. It minimizes the communication overheads by use of algebraic property of polynomial which considers the aggregation metric as number of bytes in all packets. In all clustering algorithm does well for the static conditions but needs special consideration for the changing environmental conditions. To ensure the long network life time, nodes used for the sensing purpose should be self configured. Also data aggregations protocols are required to improve the losses caused due to collision, delays from waste of bandwidths.

II. ARCHITECTURES OF DATA AGGREGATION

Based on various applications and requirements there are several existing architectures for data aggregation. They are Centralized, Decentralized, Cluster based, Tree based, Grid, Chain based architectures.

Table 3 Protocols based on different architectures

Protocol	Organization type	Objectives	Characteristics
LEACH	cluster	Network lifetime: number of nodes that are alive, latency	Randomized cluster head rotation, non-uniform energy drainage across different sensors.
HEED	cluster	Lifetime: number of rounds until the first node death	Assumption: Multiple power levels in sensors. Cluster heads are well distributed. Achieves better performance than LEACH
PEGASIS	chain	Lifetime: average energy expended by a node	Global knowledge of the network is required. Considerable energy savings compared to LEACH.
Hierarchical chain based protocols	chain	Energy× delay	Binary chain based scheme is eight times better than LEACH and the three level scheme is 5 times better than PEGASIS.
EADAT	tree	Lifetime: number of alive sensors at the end of simulation time	Sink initiated broadcasting approach. It is not clear how to choose the threshold power (P_{th}) for broadcasting help messages. No comparisons made with other existing aggregation algorithms.
PEDAP-PA	tree	Lifetime: time until the death of last node	Minimum spanning tree based approach. Achieves two times performance improvement compared to LEACH, PEGASIS.

III. PERFORMANCE METRICS

The design of Wireless Sensor Networks is a challenge because many influencing factors such as fault tolerance, scalability, production cost, operating environment, network topology, hardware constraints, transmission media, power consumption and others have to be considered.

The performance of the network is then measured based on quantifiable parameters called performance metrics.

Network Lifetime: Network lifetime is defined as the number of data aggregation rounds till x % of sensors die , where x is specified by the system designer. For instance, in applications where the time that all nodes operate together is vital, lifetime is defined as the number of rounds until the first sensor is drained of its energy.

Data accuracy: The definition of data accuracy depends on the specific application for which the sensor network is designed. For instance, in a target localization problem, the estimate of target location at the sink determines the data accuracy.

Latency: Latency is defined as the delay involved in data transmission, routing and data aggregation. It can be measured as the time delay between the data packets received at the sink and the data generated at the source nodes.

Energy Efficiency: The functionality of the sensor network should be extended as long as possible. In an ideal data aggregation scheme, each sensor should have expended the same amount of energy in each data gathering round. A data aggregation scheme is energy efficient if it maximizes the functionality of the network. If we assume that all sensors are equally important, we should minimize the energy consumption of each sensor. This idea is captured by the network lifetime which quantifies the energy efficiency of the network.

Bandwidth, Capacity and Throughput: These indicate the capacity of data which can be sent over a link within a given time, however since the data size is very small bandwidth rarely matters.

Hop Count: No of hop in communication determine the cost of path, and eventually the energy consumed in the process.

Signal Strength: SNR as an indication for the link quality and the distance between two nodes is helpful to compute and determine the nodes and their reach ability during the communication process.

IV. RELATED RESEARCH WORK

Yi *et al.* [17] proposed a HEER a Hamilton Energy-Efficient Routing Protocol. It is improved by the delay aware and energy efficient protocol. When compared with the traditional cluster based protocol it was intended for saving the energy in network admin and load balancing. In the network the Hamilton path presents in every cluster for link member and for initialization the greedy algorithm is used it transmits the data and cluster was formed. This HEER algorithm is not affected by the high delays and the node does not need any data about the world location when compared with the classic chain based protocols. PEGASIS is an example of cluster chain based protocols.

Cheng *et al.* [18] introduced a data fusion technique for wireless sensor network has a delay aware network structure. In this the nodes are formed into clusters with many sizes and communication taken place in a interleaved manner between each and every cluster having a centralized fusion. The optimization process used to find the distances of intra cluster communication. The proposed method clearly shows that in aggregation structure the data fusion keeps the consumption of energy at a low level partially and the delay get reduced in the aggregation process when compared with the existed method.

Thakkar *et al.* [19] proposed a routing algorithm EDIT Energy Delay Index for Trade Off. The objectives like energy and delay are minimized. The EDIT used to select the cluster head (CH) and next hop base on the requirements of energy and delay. There are two characteristics used in the proposed method. They are hop count, distance between the nodes and the Euclidean distance referred by sink. To get the closest data the test bed is implemented for the proceeding process. The author states that it is the first algorithm used different ways for finding the distance of delay and thus the simulation results shows the enough insights without implementing the test bed.

Cai *et al.* [20] presented a real-time routing protocol that provides a high quality of service in the rate of time delay in the industrial area for mobile wireless sensor. Firstly the location information is extracted by the geographic routing and the nodes that contain the routing table information. Then the node which lies adjacent to the end node was designed to reduce hop counts and achieve transfers of fast data package. It optimized the packets forwarding hops, which in turns minimize the time delay efficiently. The Simulation results show that the proposed algorithm was better and suitable for wireless sensor network when compared with the existed method.

Liu *et al.* [21] presented the delay in performance for wireless sensor network using a cluster tree topology. In this network it highly depends on the end to end delay on the sink and the allocation of resources in cluster head and the sensors relative location. The inter cluster traffic transmission is used to find the delay in an end to end transmission. In a particular allocation of estimated timelines the cluster head is introduced by the analytical model and the inter cluster traffic transmissions.

Yao *et al.* [22] introduced a protocol named Energy Efficient Delay Aware Lifetime (EDAL) is a protocol for balanced data collection. In wireless sensor network the EDAL was widely used for transmitting the packets. It is similar to OVR that treats packet latency to the end of delivery that perform the cost of goods delivered. Here introduced an ant colony gossiping associates with centralized heuristic search in table and distributed heuristic to reduce the computational overhead. The lifetime of system is increased by the new algorithm using load balancing for individual nodes.

We can develop an efficient optimization algorithm such as optimal data aggregation scheme. There are two different nodes involved in the wireless sensor network such as the normal sensor node and sink node also referred as aggregator node. At first, we have to compute the data aggregation node in the sensor networks based on the lifetime. The lifetime of sensor nodes depends on time varying factors, such as consumed energy for sending data from the leaves to aggregated node, queuing delay during aggregation, and tree depth delay. The time varying lifetime will be optimized using Modified Cuckoo search (MCS) Algorithm and then higher lifetime nodes are selected as aggregated nodes. Then Modified time on task (MTOT) method is used to minimize the waiting time of aggregation delay. Finally, Gravitational search algorithm (GSA) used to compute lifetime efficient routing path between source to destination via aggregated nodes.

V. EXPERIMENTAL APPORACH

Many network details in WSNs are not finalized and standardized. Building a real WSNs test is very costly. Running real experiments are always time consuming. Therefore, WSNs simulation is important for WSNs development. Protocols, schemes, even new ideas can be evaluated in a very large scale. WSNs simulators allow users to isolate different factors by tuning configurable parameters.

NS2

NS2 is the abbreviation of Network simulator version two, which first been developed by 1989 using as the REAL network simulator. NS-2 is a discrete event network simulator built in Object-Oriented extension of Tool Command Language and C++. NS-2 can support a considerable range of protocols in all layers.

TOSSIM

It is an emulator specifically designed for WSN running on TinyOS, which is an open source operating system targeting embedded operating system. TOSSIM is a bit-level discrete event network emulator built in Python, a high-level programming language emphasizing code readability, and C++. People can run TOSSIM on Linux Operating Systems or on Cygwin on Windows. TOSSIM also provides open sources and online documents.

EMSTAR

EmStar is an emulator specifically designed for WSN built in C, and it was first developed by University of California, Los Angeles. EmStar is a trace-driven emulator running in real-time. People can run this emulator on Linux operating system. This emulator supports to develop WSN application on better hardware sensors. Besides libraries, tools and services, an extension of Linux microkernel is included in EmStar emulator.

OMNET++

OMNeT++ is a discrete event network simulator built in C++. OMNeT++ provides both a noncommercial license, used at academic institutions or non-profit research organizations, and a commercial license, used at "for-profit" environments. This simulator supports module programming model. Users can run OMNeT++ simulator on Linux Operating Systems, Unix-like system and Windows. OMNeT++ is a popular non-specific network simulator, which can be used in both wire and wireless area. Most of frameworks and simulation models in OMNeT++ are open sources.

J-SIM

J-Sim is a discrete event network simulator built in Java. This simulator provides GUI library, which facilities users to model or compile the Mathematical Modeling Language, a "text-based language" written to J-Sim models. J-Sim provides open source models and online documents. This simulator is commonly used in physiology and biomedicine areas, but it also can be used in WSN simulation.

AVRORA

Avrora is a simulator specifically designed for WSNs built in Java. . Avrora provides a wide range of tools that can be used in simulating WSNs. Avrora also supports energy consumption simulation. This simulator provides open sources and online documents.

VI. CONCLUSION

This paper presented data aggregation based networks in wireless Sensor networks, their architecture and comparison. This paper also presented the various simulation tools. We have Combining aspects such as security, data latency and system lifetime in the context of data aggregation is worth exploring in future research work. A systematic study of the relation between energy efficiency and system lifetime is an avenue of future research.

REFERENCES

[1]Liu, Bei and Yuanming Wu. "A Secure and Energy-Balanced Routing Scheme For Mobile Wireless Sensor Network". Wireless Sensor Network 07.11 (2015): 137-148.

[2]Norouzi, Ali and Abdul Halim Zaim. "An Integrative Comparison of Energy Efficient Routing Protocols In Wireless Sensor Network". Wireless Sensor Network 04.03 (2012): 65-75.

[3]Shahraki, Amin, Marjan Kuchaki Rafsanjani, and Arsham Borumand Saeid. "A New Approach For Energy And Delay Trade-Off Intra-Clustering Routing In Wsns". Computers & Mathematics with Applications 62.4 (2011): 1670-1676.

[4]Ammari, Habib M. "On the Energy-Delay Trade-Off in Geographic Forwarding In Always-On Wireless Sensor Networks: A Multi-Objective Optimization Problem". Computer Networks 57.9 (2013): 1913-1935.

[5]Nguyen, Ngoc-Tu et al. "On Maximizing the Lifetime for Data Aggregation In Wireless Sensor Networks Using Virtual Data Aggregation Trees". Computer Networks 105 (2016): 99-110.

[6]Haseeb, Khalid et al. "Adaptive Energy Aware Cluster-Based Routing Protocol For Wireless Sensor Networks". Wireless Networks (2016)

[7]Mantri, Dnyaneshwar, Neeli Rashmi Prasad, and Ramjee Prasad. "Two Tier Cluster Based Data Aggregation (TTCDA) For Efficient Bandwidth Utilization In Wireless Sensor Network". Wireless Personal Communications 75.4 (2013): 2589-2606.

[8]Muthu Krishnan, A. and P. Ganesh Kumar. "An Effective Clustering Approach With Data Aggregation Using Multiple Mobile Sinks For Heterogeneous WSN". Wireless Personal Communications 90.2 (2015): 423-434.

[9]Campobello, Giuseppe, Antonino Segreto, and Salvatore Serrano. "Data Gathering Techniques For Wireless Sensor Networks: A Comparison". International Journal of Distributed Sensor Networks 2016 (2016): 1-17.

[10]Li, Wuyungerile et al. "Waterfalls Partial Aggregation In Wireless Sensor Networks". International Journal of Distributed Sensor Networks 2016 (2016): 1-11.

[11]Joo, Changhee and Ness B. Shroff. "On The Delay Performance Of In-Network Aggregation In Lossy Wireless Sensor Networks". IEEE/ACM Transactions on Networking 22.2 (2014): 662-673.

[12]Liu, Bing-Hong and Jyun-Yu Jhang. "Efficient Distributed Data Scheduling Algorithm for Data Aggregation In Wireless Sensor Networks". Computer Networks 65 (2014): 73-83.

[13]Parmar, Keyur and Devesh C. Jinwala. "Concealed Data Aggregation In Wireless Sensor Networks: A Comprehensive Survey". Computer Networks 103 (2016): 207-227.

[14]Villas, Leandro A. et al. "An Energy-Aware Spatio-Temporal Correlation Mechanism To Perform Efficient Data Collection In Wireless Sensor Networks". Computer Communications 36.9 (2013): 1054-1066.

[15]Sitarz, Piotr and Bartosz Powałka. "Modal Parameters estimation using antcolony Optimisation algorithm". (2015): 70-310.

[16]A. Villas, Leandro et al. "Dynamic and Scalable Routing To Perform Efficient Data Aggregation In Wsns". IEEE (2011)

[17]Yi, D. and H. Yang. "HEER – A Delay-Aware And Energy-Efficient Routing Protocol For Wireless Sensor Networks". Computer Networks 104 (2016): 155-173.

[18]Cheng, Chi-Tsun, Henry Leung, and Patrick Maupin. "A Delay-Aware Network Structure for Wireless Sensor Networks With In-Network Data Fusion". IEEE Sensors Journal 13.5 (2013): 1622-1631.

[19]Thakkar, Ankit and Ketan Kotecha. "Cluster Head Election For Energy And Delay Constraint Applications Of Wireless Sensor Network". IEEE Sensors Journal 14.8 (2014): 2658-2664.

[20]Cai, Hu, Yin Zhang, and Hehua Yan. "A Delay-Aware Wireless Sensor Network Routing Protocol For Industrial Applications". (2016)

[21]Liu, Wenjuan, Dongmei Zhao, and Gang Zhu. "End-To-End Delay And Packet Drop Rate Performance For A Wireless Sensor Network With A Cluster-Tree Topology". Wireless Communications and Mobile Computing 14.7 (2012): 729-744. Web. 1 Dec. 2016.

[22]Yao, Yanjun, Qing Cao, and Athanasios V. Vasilakos. "EDAL: An Energy-Efficient, Delay-Aware, And Lifetime-Balancing Data Collection Protocol For Heterogeneous Wireless Sensor Networks". IEEE/ACM Transactions on Networking 23.3 (2015): 810-823.