ISSN (Online): 2230-7893
www.IJCEM.org

Analysis of Data Aggregation Techniques in Wireless Sensor Networks

Prakashgoud Patil[1], Umakant Kulkarni [2]

[1]B.V.B. College of Engineering and Technology, Hubli, Karnataka, INDIA

[2]S.D.M. College of Engineering and Technology, Dharwad, Karntaka, INDIA

Abstract

The main goal of data aggregation algorithms is to gather and aggregate data in Wireless Sensor Network (WSN) in an energy efficient manner so that network lifetime is enhanced. Data aggregation may be effective technique in this context because it reduces the number of packets to be sent to sink by aggregating the similar packets. The performance of data aggregation protocols are characterized by performance measure such as energy consumption, latency ad data accuracy. In this paper, we present a survey of data aggregation algorithms and approaches for handling the trade-offs in data aggregation schemes.

Keywords: *Wireless Sensor Network, Data Aggregation, Clustering and Energy Efficiency.*

1. Introduction

In typical wireless sensor networks, sensor nodes are usually resource-constrained and battery-limited. In-network aggregation is the global process of gathering and routing information through a multi hop network, processing data at intermediate nodes with the objective of reducing resource consumption (in particular energy), thereby increasing network lifetime. In order to save resources and energy, data must be aggregated to avoid overwhelming amounts of traffic in the network. The aim of data aggregation is that eliminates redundant data transmission and enhances the lifetime of energy in wireless sensor network.

2. Preliminaries of Data Aggregation

Data aggregation requires a different forwarding paradigm compared to classic routing. Classic routing protocols typically forward data along the shortest path to the destination (with respect to some specified metric). If, however, we are interested in aggregating data to minimize energy expenditure, nodes should route packets based on the packet content and choose the next hop in order to promote in-network aggregation. This type of data forwarding is often referred to as data centric routing. According to the data centric paradigm, as a node searches for the relay nodes, it needs to use metrics which take into account the positions of the most suitable aggregation points, the data type, the priority of the information, etc. In particular, the best strategy at a given node is not always to send data as soon as it is available. Waiting for information from neighboring nodes may lead to better data aggregation opportunities and, in turn, improved performance.

The performance measures of data aggregation algorithms such Network Lifetime, Latency, and Data Accuracy are described below.

Network lifetime: The network lifetime is defining the number of data fusion rounds. Till the specified percentage of the total nodes dies and the percentage depend on the application.

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 packet received at the sink and data generated at the source node.

Data accuracy: It is evaluate of ratio of total number of reading received at the base station (sink) to the total number of generated

3. Data Aggregation Techniques based on Network Architecture

Performance of data aggregation protocols are greatly affected by the network architecture as shown in the figure 1. The data aggregation based on architecture divided into parts such as: Flat Networks and Hierarchical. Hierarchical data aggregation can be further divided into four parts cluster, chain, tree and grid.

In flat network, aggregation is done in data centric routing method, the sink transmits a query message to the sensors, and sensors which have data matching the query send response messages back to the sink. Excessive communication and computation are performed in sink node, resulting in a faster depletion of its battery power. The failure of the sink node breaks down the functionality of the network. In this paper we describe the different data-aggregation protocols and highlight their advantages and limitations.

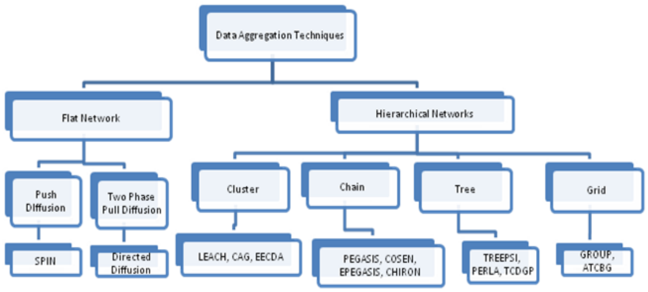

Figure 1 WSN-Network Architecture

3.1 Cluster-Based Data Aggregation Technique

Cluster-based schemes [2] also consist of a hierarchical organization of the network. However, here nodes are subdivided into clusters. Moreover, special nodes, referred to as cluster-heads, are elected in order to aggregate data locally and transmit the result of such an aggregation to the sink. In this section we analyze three cluster based protocols, such as LEACH, CAG and EECA and their pros and cons in Table 1.

3.1.1 Low-Energy Adaptive Clustering Hierarchy (LEACH)

Heinzelman et al proposed the LEACH protocol [2]. It is the first dynamic cluster head protocol specifically for WSN using homogeneous stationary sensor nodes randomly deployed. LEACH is suited for applications which involve constant monitoring and periodic data reporting. LEACH protocol runs in many rounds. Each round contains two phases: cluster setup phase and steady phase.

In cluster setup phase, it performs organization of cluster and selection of cluster head. Selected cluster heads broadcast a message to all the other sensors in the network informing that they are the new cluster heads. All non cluster head nodes which receive this advertisement decide which cluster they belong to based on the signal strength of the message received. All non-cluster head nodes transmit their data to the cluster head, while the transmits the data to the remote base station(BS). Cluster head node is much more energy intensive than being a non cluster head node. Head nodes would quickly use up their limited energy. Thus, LEACH incorporates randomized rotation of the high-energy cluster head position among the sensors.

3.1.2 Clustered Aggregation Technique (CAG)

It is an algorithm to compute approximate answers to queries by using spatial and temporal properties of data [3]. CAG forms clusters of nodes sensing similar values. It ignores redundant data using the spatial and temporal correlations provide significant energy savings. CAG can work in two modes: a) interactive mode and b) Streaming mode. CAG generates a single set of responses for a query in the interactive mode. In the streaming mode, periodic responses are generated in response to a query.

The interactive mode of CAG exploits only the spatial correlation of sensed data. CAG builds a forwarding tree when a query is sent out. Thus, the forwarding path is set along the reverse direction of the query propagation. However, the interactive mode requires the overhead for broadcasting a query each time a user wants new data from the network.

In the streaming mode of CAG takes advantage of both spatial and temporal correlations of data. A query for the streaming mode uses the clause "epoch duration i" to define the sampling frequency. The query is injected into the network only once with this clause, it generates a query reply for every i seconds.

3.1.3 Energy Efficient Clustering and Data Aggregation (EECDA)

EECDA combines energy efficient cluster based routing and data aggregation for improving the performance in terms of lifetime and stability [4]. It is for the heterogeneous WSN. EECDA balances the energy consumption and prolongs the network lifetime by a factor of 51%, when compared with LEACH.

Table 1. Comparison of Cluster Based Routing Protocols

Protocol	Advantage	Limitation
LEACH	- Randomized Cluster Head Selection. - Improves the system performance lifetime and data accuracy of the network.	-Cost to form a cluster is expensive. -CH will be concentrated on one part of the network and clustering terminates in a constant number of iterations.
CAG	-It provides energy efficient aggregation results with small and negligible error.	-It's save energy when only few nodes change clusters. -Resilient to the packet loss.
EECDA	-It improves the N/W performance by using some heterogeneous node in the network.	The election process of CHs makes the network unstable.

3.2 Chain Based Data Aggregation Techniques

In the cluster based network, if the cluster heads are far away from the sink then it need excessive energy to communicate the sink. In chain based data aggregation the data is sent only to the closest neighbor. This section includes PEGASIS, COSEN, Enhanced PEGASIS, CHIRON protocols and point out their pros and cons. These protocols are compared in Table 2.

3.2.1 Power Efficient GAthering in Sensor Information System (PEGASIS)

It is a near optimal chain based power efficient protocol based on LEACH. The cluster formation and cluster head selection is not used in PEGASIS[5]. Each node determines the distance to its neighbors using the signal strength and then adjusts the signal strength to communicate only with the closest neighbor. Collected data moves across the nodes, gets aggregated at each node, and eventually, a single designated node transmits data to the base station. Nodes take turns in transmitting to the base station so that the power dissipation for communicating with the base station is distributed uniformly among all the nodes.

In PEGASIS the chain construction is done in greedy fashion with the assumption that all the nodes have global knowledge of the network. The leader in each round of communication is selected from a random location in the chain.

3.2.2 Chain Oriented Sensor Network for Efficient Data Collection (COSEN)

It is a two-tier hierarchical chain-based routing scheme. COSEN compared to PEGASIS, it can alleviate the transmission delay and energy consumption. In that scheme, sensor nodes are geographically grouped into several low-level chains [6]. For each low-level chain, the sensor node with the maximum residual energy is elected as the chain leader. Moreover, with the low-level leaders, a high-level chain and its corresponding chain leader will be eventually formulated. In data communications, all common (normal) nodes perform a similar procedure as that in PEGASIS to send their fused data, via their respective low-level leaders and the high-level leader, toward the BS. But COSEN introduce a lot of redundant transmission paths, especially for those nodes which are nearest to the BS.

3.2.3 Enhanced PEGASIS

In this method, the sensing area, centered at the BS, is circularized into several concentric cluster levels. For each cluster level, based on the greedy algorithm of PEGASIS, a node chain is constructed. In data transmission, the common nodes transfer their sensing data to its chain leader in a similar way as the PEGASIS. After that, from the highest (farthest) cluster level to the lowest (near to the BS), a multi-hop and leader-by-leader data propagation task will be followed.

3.2.4 Chain-Based Hierarchical Routing Protocol (CHIRON)

It is an energy-efficient hierarchical chain-based routing protocol, named as CHIRON, for wireless sensor networks. In CHIRON, divides the sensing area into several fan-shaped groups. The sensor nodes within each group are self organized into a chain. It considers the node with a maximum residual energy as chain leader candidate. For avoiding a longer transmission would be incurred among chain leaders, the nearest downstream chain leader will be elected for relaying the aggregated sensing information[7].

Table 2. Comparison of Chain Based Routing Protocols

Table2. Chain Based Data Aggregation Techniques		
Protocol	**Advantage**	**Limitation**
PEGASIS	-Balance network energy dissipation. -Considerable energy savings compared to LEACH	-Communication delay can be large due to long single chain -Delay might be intolerable in large Network
COSEN	-Alleviate the transmission delay and energy consumption	-Lot of redundant transmission paths
Enhanced PEGASIS	-Improve the redundant transmission path and the network lifetime	-Longer transmission delay, consume more energy Node's residual energy is not considered for leader node election
CHIRON	-Reduces data propagation delay and redundant transmission Path to save energy	-Multiple short chains

3.3 Tree based Data Aggregation Techniques

In this type of scheme nodes are organized in a tree topology where the sink is represented as a root. All the intermediate nodes perform the aggregation and transfer it to the root (sink). Energy efficient tree construction is the main aspect in the tree based approach[11,12]. This section includes TREEPSI, PERLA and TCDGP protocols and with their benefits and limitations.

3.3.1 Tree-based Efficient Protocol for Sensor Information (TREEPSI)

It is a tree based protocol. Before data transmission phase, it will select a root node among the sensor nodes. There are two ways to build the tree path first the computing the path centrally using the sink and broadcasting the path information to the network, second can be a common algorithm in each node. At the initial phase, the root is identified by id = j. Root will create data gathering process from the children nodes using any standard tree traversal algorithm. Then perform the data transmission phase after building the tree. All the leaf nodes will start sending the sensed data towards their parent nodes. The parent nodes will collect the received data together with their own data that is then sent to their parents. The process will repeat until the root node has no more data to send. After the root node has aggregated the data, it sends the collected data directly to the sink. The WSN will then re-select a new root node. The new root identification number would be j + 1. The initial phase is then repeated and the tree path will not change until the root node is dead[8].

3.3.2 Power Efficient Routing with Limited Latency (PERLA)

It is to enhance the robustness in the network. It considers the link failure and node failure problems. PERLA recognize and recover from link failures by using some specific procedure. PERLA identifies link failures by using acknowledgments. In the initialization phase, each node selects a primary parent for packet forwarding, and a set of upper-layer neighbor nodes as backup parents. With respect to pure multi-path approaches, PERLA reduces the overall traffic in the network by performing retransmissions only if necessary. It does not rely on link-layer acknowledgments and retransmissions.

3.3.3 Tree-Clustered Data Gathering Protocol (TCDGP)

It is to combine cluster-based and tree-based protocols. It performs data collection in three phases. Such as a) Cluster Establishment phase. b) Constructing Cluster Based Tree phase and c) Data aggregation phase [10].
Cluster Establishment phase consists of two major steps: cluster formation and cluster head selection. The base station forms the clusters and selects the cluster head; it may be different in each round. During the first round, the base station first splits the network into two sub clusters, and proceeds further by splitting the sub clusters into smaller clusters. Base station repeats the cluster splitting process until the desired number of clusters is attained. When the splitting algorithm is completed, the base station will select a cluster head for each cluster according to the location information of the nodes. For a node to be a cluster head, it has to locate at the center of a cluster.
Once a node is selected to be a cluster head, it broadcasts a message in the network and invites the other nodes to join its cluster. The other nodes will choose their own cluster heads and send join messages according to the power of the many received broadcast messages.

In data aggregation phase after the routing mechanism has established, nodes transmits their gathered data to higher level nodes. Then the higher level nodes will fuse the received data and send it to next level nodes. This process will be repeated until aggregated data reaches the root node. The above discussed tree based routing protocols comparisons are given in Table 3.

Table 3- Tree Based Aggregatio Techniques

Protocol	Advantage	Limitation
TREEPSI	- Power consumption is less in data transmission	-The path has made a detour in the topology
PERLA	-Avoids unnecessary route changes.	-Needs more energy for the error detection and the recovery procedures.
TCDGP	-Reduce energy consumption	-Need methods for recovery procedures.

3.4 GRID based Data Aggregation Techniques

In grid-based data aggregation, the data aggregator is fixed in each grid and it aggregates the data from all the sensors. Hence, the sensors within a grid do not communicate with each other. Any sensor node within a grid can assume the role of a data aggregator in terms of rounds until the last node dies. It is best suitable for military surveillance and weather forecasting. In this section GROUP and ATCBG protocols are discussed and its comparisons are in Table 4.

3.4 GROUP

It is an energy-efficient and cluster-based routing protocol for wireless sensor networks. In GROUP[9], the nodes are organized into clusters. One node is selected as the cluster head (CH) in each cluster. And all cluster heads form a virtual cluster grid. The data queries will be transmitted from sinks to all nodes via cluster heads. And the data matched the query are routed back to sinks via cluster heads. GROUP select cluster heads dynamically. The data forwarding can be done in three processes the cluster grid construction process, query forwarding and data forwarding.
In Cluster grid construction phase after the wireless sensor network is deployed, all sinks in the network will elect one

sink as the primary sink (PS), which initiates the cluster grid construction process, based on their location. The PS is closer to the center of network than other sinks in order to keep a minimum duration of grid construction.
Query forwarding phase queries are forwarded through limited- broadcast and unicast respectively. There are two typical classes of queries sent by sinks, i.e. location unaware query and location-aware query. In GROUP, The location-unaware query is transmitted from one of the sink to its closest cluster head. The location-aware query will forward the query to one of its downstream cluster heads which is closest to the destination area mentioned in the query.

In Data forwarding phase a sensor node receives the query from its cluster head; it will check the query and collect the data. If the collected data match the query, it sends out data to its cluster head through short-range radio. The data packet will be forwarded recursively by the cluster head to its upstream cluster head till it reaches the sink which generated the query. In GROUP, cluster heads can perform data aggregation in order to reduce the number of data packets transmission.

3.5 Aggregation Tree Construction Based On Grid (ATCBG)

The aggregation tree construction algorithm (ATCBG) is to having some improvements over GROUP. The main idea of ATCBG is that aggregation tree is constructed by taking the sink as the center of a grid. The whole network is divided into grids. Each grid forms a cluster. The cluster head is elected by considering residual energy, distance to the center of the grid and other factors. The cluster head take responsible for data aggregation. All the cluster heads form a tree-structure.
The aggregation tree construction is initiated by sink. Sink broadcasts tree construction message. In Cluster Head Replacing Scheme: The cluster head will consume more energy due to receiving and fusing all the data from its member nodes and child nodes. To avoid premature death of the cluster head, the cluster head must be replaced after a certain time.
ATCBG replaces the cluster head when its residual energy is below half of the energy with which it was electing for cluster head. When residual energy below the threshold, the cluster head sends replacing cluster head message for replacing the cluster head.

In Data Transmission the cluster member nodes first send the collected data to the corresponding cluster head. The cluster head fuses the data after receiving all the data from member nodes and its child nodes. And then cluster head sends the fused data to its parent. The process continues until the data is sent to the sink.

Table 4. GRID based Techniques

Protocol	Advantage	Limitation
GROUP	Distribute the energy load among sensor in the n/w and provide in-network processing.	1. Cluster head selection considers only the distance of the grid. 2 aggregation tree reconstructed periodically
ATCBG	1. Cluster head selection considers distance and energy. 2 Cluster head energy below half of the energy will be replaced.	Tree construction based on energy only

4. Conclusions

The comprehensive survey data-aggregation algorithms in Data Aggregation in Wireless Sensor Network is studied. All of them focus on optimizing important performance measures such as network lifetime; data latency, data accuracy, and energy consumption. We have described the main features, the advantages and disadvantages of various data aggregation algorithm. However, the performance of the data aggregation protocol is strongly coupled with the infrastructure of the network. Although, many of the data-aggregation techniques discussed look promising, there is significant scope for future research. Combining aspects such as security, data latency, and system lifetime in the context of data aggregation is worth exploring.

References

[1]. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, A survey on Sensor Networks, IEEE Communications Magazine, vol. 40, Issue: 8, pp. 102-114, August 2002.
[2]. W. Heinzelman, A. Chandrakasan, and H. balakrishnan, "An Energy-Efficient Communication Protocol for Wireless Microsensor Networks, Proc. Hawaii Conf. System Sciences, Jan. 2000.
[3]. SunHee Yoon and Cyrus Shahabi, March 2007, "The Clustered AGgregation (CAG) Technique Leveraging Spatial and Temporal Correlations in Wireless Sensor Networks", ACM Transactions on Sensor Networks (TOSN), Vol. 3, Issue 1, No.3
[4]. D. Kumar, T.C. Aseri, R.B. Patel "EECDA: Energy Efficient Clustering and Data Aggregation Protocol for Heterogeneous Wireless Sensor Networks " Int. J. of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844 Vol. VI (2011), No. 1 (March), pp. 113-124
[5]. Lindsey S Raghavendra C, "PEGASIS Power- efficient Gathering in Sensor Information System," In Proceeding of

the IEEE Aerospace Conference. Montana IEEE Aerospace and Electronic Systems Society 2002, pp.1125~1130.

[6]. N. Tabassum, Q. E. K. M. Mamun, and Q. Urano, “COSEN: A Chain Oriented Sensor Network for Efficient Data Collection,” Proceedings of the Global Telecommunications conference,Vol. 6, pp. 3525-3530, 2003

[7]. Kuong-Ho Chen , Jyh-Ming Huang, Chieh- Chuan Hsiao CHIRON: An Energy-Efficient Chain-Based Hierarchical Routing Protocol in Wireless Sensor Networks, IEEE. 2009

[8]. S.S. Satapathy and N. Sarma , TREEPSI: tree based energy efficient protocol for sensor information ,Wireless and Optical Communications Networks 2006, IFIP International Conference 11-13 April 2006 .

[9]. Liyang Yu, Neng Wang, Wei Zhang, Chunlei Zheng, “GROUP: a Grid-clustering Routing Protocol for Wireless Sensor Networks,” In proceedings of 2006 IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2006), Wuhan, China, 2006, pp.1~5

[10]. Gurpreet Singh Chhabra, Dipesh Sharma, “Cluster-Tree based Data Gathering in Wireless Sensor Network “International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231- 2307, Volume-1, Issue-1, March 2011

[11]. Jian Shu, Yebin Chen,Linlan Liu, Sheng Zhang and Jun Li Study on Aggregation Tree Construction Based on Grid journal of computers, vol. 5, no. 6, june 2010

[12]. [8] Messina, D.; Ortolani, M.; Lo Re, G., Oct. 2007 , “A Network Protocol to Enhance Robustness in Tree-Based WSNs Using Data Aggregation”, I EEE Internatonal Conference on Mobile Adhoc and Sensor Systems, MASS 2007, Issue , 8-11 Page(s):1-4.

[13]. Q. Jiang and D. Manivannan , Routing Protocols for Sensor Networks, IEEE Consumer communications and Networking Conference, 2004.

[14]. S. Madden, M. Franklin, J. Hellerstein, and W. Hong, “TAG: a Tiny AGgregation service for ad- hoc sensor networks,” In 5th Annual Symposium on Operating Systems Design and Implementation (OSDI), December 2002, Pages: 131-146.

[15]. W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “An Application-Specific Protocol Architecture for Wireless Microsensor Networks,” IEEE Trans Wireless Commun., Oct. 2002, pp. 660–70.

[16]. Ramesh Rajagopalan and Pramod K. Varshney, 2006, "Data aggregation techniques in sensor networks: A survey", IEEE Communications Surveys and Tutorials, vol. 8, no. 4, 4th quarter.

Prakashgoud Patil received his B.E Degree in Computer Science and Engineering from Karnataka University, Dharwad (India) in 1991 and ME degree in Computer Science from Thapar Institute of Technology (TIET), Patiala, Punjab. Presently, he is perusing Ph.D degree in Computer Research Center at S.D.M Engineering College, Dharwad affiliated to Visvesvaraya Technological University, Belgaum under the guidance of Dr. Umakant P. Kulkarni.

He is currently working as Assistant Professor in Master of Computer Applications Department, B.V.B College of Engineering and Technology, Hubli, Karantak (India). He has presented and published research papers at national and IEEE International conferences and won best paper awards. He is also actively involved in the development of business applications using open source software & technologies and his research interests include Computer Networking, Wireless Sensor Networks and Cyber Security.

Prof. Umakant P. Kulkarni received his PhD degree from Shivaji University, Kolhapur. Research Center:- Walchand College of Engineering & Tech, Sangli, Maharashtra in Nov 2007. He received his Masters degree (M.E) from PSG, Coimbatore. He is currently working as Professor in Computer Science Department, SDMCET, Dharwad, Karnataka. He has published and presented a number of papers in many reputed journals and at International IEEE conferences as well. This apart he also has presented technical talks and tutorials on various aspects of Mobile Agents. His research interests include mobile agents, computer networking, and distributed networking.