Research Article *ISSN: 2319-507X*
Dipak Khadse, IJPRET, 2013; Volume 1(8): 23-34 *IJPRET*

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

COMPARATIVE ANALYSIS OF VARIOUS DATA AGGREGATION TECHNIQUES IN WSN

DIPAK B. KHADSE, MANISH NARNAWARE

Department of Computer Science & Engineering, G. H. Raisoni College of Engineering, Nagpur.

Accepted Date:

27/02/2013

Publish Date:

01/04/2013

Keywords

Cluster head,

Data aggregation,

Redundant,

Wireless sensor networks,

Aggregation architecture,

LEACH,

PEGASIS

Corresponding Author

Mr. Dipak B. Khadse

Abstract

Sensor networks consist of several sensor nodes which co-operatively send sensed data to base station. Data aggregation is essential for the efficient operation of wireless sensor networks. Data Aggregation has been widely recognized as an efficient method to reduce energy consumption by reducing number of packet sent. The main goal of data aggregation is to gather and aggregate data in energy efficient manner so that network lifetime is enhanced. In this paper we present comparative analysis of various existing data aggregation techniques in wireless sensor network. Our aim is to provide a good understanding of data aggregation in WSN and its related issue which propose directions for future research in this area.

1. INTRODUCTION

The WSN consist of hundred to thousand of inexpensive wireless nodes in fig.1. Each with some computational power and sensing capability, operating in an unattended mode. Sensor nodes have limited processing capability and low battery power. It has also a sensing element and transreciever. Sensor nodes sense the physical environment and send the data in the term of signal to base station. The sensor nodes usually scattered in sensor field as shown in figure. Each of these scattered nodes has capabilities to collect data and route data back to the sink. While sending data by its transreciever some amount energy required, so energy conservation is the important factor in sensor network.

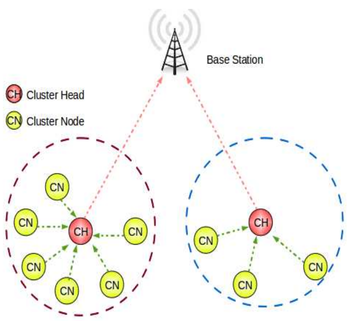

Figure 1 Wireless Sensor Network

Data aggregation [1] is good technique to save the precious energy of sensor nodes. Usually in sensor network thousand of sensor nodes deployed for area monitoring. Most of them sense the environment and send the data to the base station and at the base station we have to combine all the information for desired output. If we aggregate the data before reaching the base station we have to combine all the information for desired ouput. If we aggregate the data before reaching the base station we can potentially decrease number of packets in the network so we will have to send less number packets to base station and that can save the energy of sensor nodes. These types of data aggregation are called In-Network data aggregation where packets are combined before reaching the base station. The main idea of the data aggregation and in-network processing approaches is to combine the data arriving from different sensor nodes at certain aggregation points eliminates redundancies by performing simple processing at the aggregation points, and minimize the total amount of data transmission before forwarding data to base station. The main goal of data aggregation is a gather data and aggregate data in efficient manner so that network lifetime is enhanced.

The remainder of the paper is organized as follow. Section 2 introduces the data aggregation in WSN .Section 3 Introduce different data aggregation techniques.

2. DATA AGGREGATION

Data Aggregation [2] is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide fused information to the based station. Data latency and accuracy are important in much application such as environment monitoring, where the freshness of data is also an important factor. It is critical to developed energy-efficient and fast data aggregation algorithm. There are several factors which determine the energy efficiency of a sensor network, such as network architecture, the data-aggregation mechanism, and the underlying routing protocol.

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.

Network lifetime, data accuracy, and latency are some of the important performance measures of data-aggregation algorithms.

Network Lifetime: Network lifetime is defined as the number of rounds until the first sensor is drained of its energy. The main idea is to perform data aggregation such that there is uniform energy drainage in the network. In addition, energy efficiency and network lifetime are synonymous in that improving energy efficiency enhances the lifetime of the network.

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 the 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.

2.1 Objective of Data Aggregation

Main Objective of Data aggregations are to reduce energy consumption by reducing redundant packet and increase network life time.

3. DATA AGGREGATION BASED ARCHITECTURE

The architecture [3, 4] of the sensor network plays a vital role in the performance of different data-aggregation protocols. In this section we survey several data-aggregation protocols which have specifically been designed for different network architectures.

1. Flat network

2. Hierarchical network

Flat Network

In flat networks, each sensor node plays the same role and is equipped with approximately the same battery power. In such networks, data aggregation is accomplished by data centric routing where the sink usually transmits a query message to the sensors, for example, via flooding and sensors which have data matching the query send response messages back to the sink.

Hierarchical Networks

A flat network can result in excessive communication and computation burdens at the sink node, resulting in a faster depletion of its battery power. The death of the sink node breaks down the functionality of the network. Hence, in view of scalability and energy efficiency, several hierarchical data-aggregation approaches have been proposed. Hierarchical data aggregation involves data fusion at special nodes, which reduces the number of messages transmitted to the sink. This improves the energy efficiency of the Network.

4. PROTOCOL BASED ON NETWORK ARCHITECTURE

The topology of the sensor network plays a vital role in the performance of different data aggregation protocols. In this section, we explore data aggregation protocols which have specifically been designed for different network topology

4.1 Data Aggregation in Flat Network Topology

In flat network topology [4] each sensor node plays the same role and is outfitted with roughly the same running power. In such networks, data aggregation is accomplished by data.

Hierarchical network	Flat Network
Data aggregation performed by cluster heads or leader node Overhead involved in cluster or chain formation throughout the network	Data aggregation is performed by different nodes along the multi-hop path 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 1 Hierarchical Network vs. Flat Network

A. Flooding and gossiping

Flooding and gossiping [5] are two conventional mechanisms to relay data in sensor networks without the need for any routing algorithms and topology maintenance. In flooding, each sensor receiving a data packet broadcasts it to all of its neighbours and this process continues until the packet arrives at the destination or the maximum number of hops for the packet is reached. On the other hand, gossiping is a slightly enhanced version of flooding where the receiving node sends the packet to a randomly selected neighbour, which picks another random neighbour to forward the packet to and so on.

B.Directed diffusion

Directed diffusion (DD) [6] is a popular data aggregation paradigm for wireless sensor networks. It is a data-centric and application aware paradigm, in the sense that all data generated by sensor nodes is named by attribute-value pairs. Such a scheme combines the data coming from different sources enroute to the sink by eliminating redundancy and minimizing the number of transmissions. In this way, it saves the energy consumption and increases the network lifetime of WSNs. In directed diffusion, the base station requests data by broadcasting interests, The interest is defined using a list of attribute value pairs such as name of objects, interval, duration and geographical area. Each node receiving the interest can cache it for later use. As the interest is broadcasted through the network hop-by-hop, gradients are setup to draw data satisfying the query towards the requesting node. A gradient is a reply link to the neighbour from which the interest was received. It contains the information derived from the received interest's fields, such as the data rate, duration and expiration time. Each sensor that receives the interest sets up a gradient toward the sensor nodes from which it received the interest. This process continues until gradients are setup from the sources all the way back to the base station. In this way, several paths can be established, so that one of them is selected by reinforcement. The sink resends the original interest message through the selected path with a smaller interval, hence reinforcing the source node on that path to send data more frequently.

4.2 Data aggregation in Hierarchical network topology

A flat network can result in excessive communication and computation burdens at the sink node, resulting in a faster depletion of its running power. In order to deal with this issue, hierarchical network topology has been proposed. In involves data fusion at special nodes, which reduces the number of messages transmitted to the sink. This improves the energy efficiency of the network. Cluster, tree and grid networks come under hierarchical topology, which we discuss in the following section.

4.2.1 Data Aggregation in Cluster-Based Network

In energy-constrained sensor networks of large size, it is inefficient for sensors to transmit the data directly to the sink. In such scenarios, sensors can transmit data to a local aggregator or cluster head which aggregates data from all the sensors in its cluster and transmits the concise digest to the sink. This results insignificant energy savings for energy constrained sensors. In this section we discuss some clustering protocols like low Energy Adaptive Clustering Hierarchy (LEACH).

A. LEACH

Heinzelman propose the LEACH [6] protocols. It is First dynamic cluster head protocol specifically design for WSN.LEACH is suited for application which involve constant monitoring and periodic data reporting. LEACH protocol run 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. The data collection in the cluster is centralized and it is performed periodically using a TDMA (Time division multiple access) schedule created by every CH (cluster head). The sensor nodes send data to the CH according to the schedule in steady phase.

LEACH improves the system performance lifetime and data accuracy of the network but the protocol has some limitations such as that the elected CH will be concentrated on one part of the network and clustering terminates in a constant number of iterations.

B.Clustered Aggregation technique (CAG)

It is an algorithm to compute approximate answers to queries by using spatial and temporal properties of data. The main difference between LEACH and CAG is that LEACH does not provide a mechanism to compute aggregate using cluster head values.

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. The frequent rebuilding of the tree can be wasteful if the sensed data is almost the same over time. If the data is unchanged, cluster head nodes and the forwarding tree are likely to be the same.

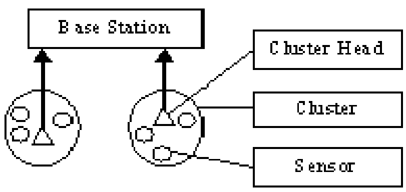

Figure 2 Cluster Based Data Aggregation

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. The CAG algorithm operates in two phases: query and response. The query phase of the CAG algorithm in the streaming mode is identical to that of the clustering algorithm in the interactive mode. The response phase algorithm is the major difference between the streaming mode and the interactive mode.

4.2.2 Chain based Data Aggregation techniques

In the cluster based network [4] all the cluster members send the data to its cluster head. Cluster heads transfer the aggregated data to the sink. If the cluster heads are far away from the sink then it need excessive energy to communicate the sink. But in chain based data aggregation the data is sent only to the closest neighbor.

In figure 3 chain based cluster sensor network the chains of nodes are constructed for data transmission to the cluster-head. All nodes in a cluster send the sensed data to their neighbor node instead of the cluster-head, while each node aggregates the data to reduce the amount of data transferred.

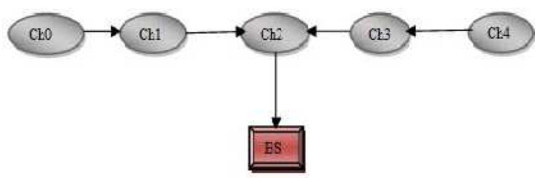

Figure 3 Chain in PEGASIS

A. 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 [1].

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. This causes nodes to die at random positions in the chain which is important to make the sensor network robust to failure. In each round, a simple control token passing is initiated by the leader to start the data transmission form the ends of the chain. The cost is very small due to the small token size.

But, the limitation of this protocol is the communication delay can be large due to long single chain. When the network size is relatively large, the delay might be intolerable. Also, the inter-node distance gets larger as the network size grows, which increases energy consumption.

4.2.3 Tree based Data Aggregation Techniques

In which all nodes are organized in form of tree means hierarchical [10][11], with the help of intermediate node we can perform data aggregation process and data transmit leaf node root node. Tree based data aggregation is suitable for applications which involve in-network data aggregation. An example application is radiation-level monitoring in a nuclear plant where the maximum value provides the most useful information for the safety of the plant. One of the main aspects of tree-based networks is the construction of an energy efficient data-aggregation tree.

A. Tree-based Efficient Protocol for Sensor Information (TREEPSI)

In [5], authors proposed a Tree-based Efficient Protocol for Sensor Information (TREEPSI). TREEPSI is tree-based protocol. In this protocol, WSNs will select a root node in all the sensor nodes; there are two ways to build the tree path. One is computing the path centrally by sink and broadcasting the path information to network.

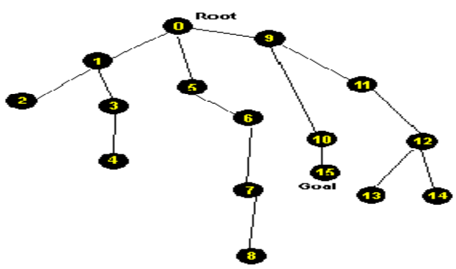

Figure 4 Tree-based networks

The other can be the same tree structure locally by using a common algorithm in each node. Initially root creates the data gathering process to their child node. The next step is data gathering after forming a tree. All the leaf nodes will start sending the sensed data towards their parent nodes. The parent nodes will collect the received data with their own data. Then send the collected data to their parent. The transmission process will be repeated until all data the received by the root node [5]. The data aggregation will be takes place at the root node, after aggregating the data, it send collected data directly to the sink node. The process will go around until the root node dead. WSN will re-select a new root node and initial phase is repeated like above. The tree path will not change until the root node dead. The length of path form end leaf node to root/chain node in TREEPSI is shortest. The data will not send for a longer path. For this reason, TREEPSI can reduce power consumption less in data transmission than other existing protocol like PEGASIS [1].

5. CONCLUSION

In this paper we have presented a various data-aggregation algorithms in wireless sensor networks. We also present the comprehensive study of various data aggregation protocols under the network architecture. Many solutions are proposed in the tree-based and cluster-based categories, but very few studies use the multipath and hybrid approaches. This leaves room for further work in this area.

6. REFERENCES

1. Sushruta Mishra, Hiren Thakkar "Features of WSN and Data Aggregation techniques in WSN: A Survey", International Journal of Engineering and Innovative Technology (IJEIT) Volume 1, Issue 4, April 2012

2. Gurpreet Singh Chhabra, Dipesh Sharma "Cluster-Tree based Data Gathering in Wireless Sensor Network", International Journal of Soft Computing and Engineering (IJSCE) Volume-1, Issue-1, March 2011.

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

4. Kavi Khedo, Rubeena Doomun, Sonum Aucharuz , "READA: Redundancy Elimination for Accurate Data Aggregation in Wireless Sensor Networks" Wireless Sensor Network, 2010, 2, 300-308, SciRes.

5. E. Fasolo et al. In-network aggregation techniques for wireless sensor networks: a survey. Trans. Wireless Communication. 14(2):70–87, Apr. 2007.

6. R. Rajagopalan and P. K. Varshney, "Data-Aggregation Techniques in Sensor Networks: A Survey," IEEE Com-munication Surveys and Tutorials, Vol. 8, No. 4, December 2006, pp. 48-63.

7. 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.

8. 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.

9. K. Akkya and M. A. Younis, "Survey of routing protocols in wireless sensor

networks", Elsevier Ad Hoc Network Journal, Vol 3/3, pp 325-349, 2005.

10. Schurgers C. and Srivastava M.B. (2001) Energy Efficient Routing in Wireless Sensor Networks. In MILCOM Proceedings on Communications for Network-Centric Operations: Creating the Information Force, McLean, VA.

11. C. Shen, C. Srisathapornphat, and C. Jaikaeo, "Sensor InformationNetworking Architecture and Applications," IEEE Pers. Commun., Aug. 2001, pp. 52–59.