

THE IMPACT OF DATA-AGGREGATION TECHNIQUES IN WSN

Dr.Nagendra Nath Giri¹,

Asso. Professor, Dept. of CS, Maharani's Science college for women ,JLB Road ,
Mysore ,Karnataka ,India,nagengiri123@gmail.com

Dr. Sumati Behera²

Asst. Professor, Dept. of CS,Maharani's Science college for women ,JLB Road ,
Mysore ,Karnataka ,India,neenagiri123@gmail.com

Abstract— Wireless sensor networks (WSNs) consist of sensor nodes. These networks have huge application in habitat monitoring, disaster management, security and military etc. Wireless sensor nodes are very small in size and have limited processing capability with very low battery power. This restriction of low battery power makes the sensor network prone to failure. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. In this paper we discuss about data aggregation and its impact in WSN

Keywords—Data aggregation, energy efficient, Impact, Wireless sensor network.

I. INTRODUCTION

Data gathering is defined as the systematic collection of sensed data from multiple sensors to be eventually transmitted to the base station for processing. Since sensor nodes are energy constrained, it is not necessary for all the sensor nodes to transmit the data directly to the base station instead data aggregation and multihop communication can be used. Usually data generated from neighboring sensor nodes is often redundant and highly correlated that depends on the phenomenon to be monitored and its precision. In addition, the amount of data generated in large sensor networks is usually enormous for the base station to process. Hence, It need methods for combining data into high-quality information at intermediate nodes which can reduce the number of packets transmitted to the base station resulting in conservation of energy and bandwidth. This can be accomplished by data aggregation. Data aggregation usually involves the aggregation of data from multiple sensor nodes at intermediate nodes and transmission of the aggregated data to the sink .The taxonomy is presented in Figure 1.0.

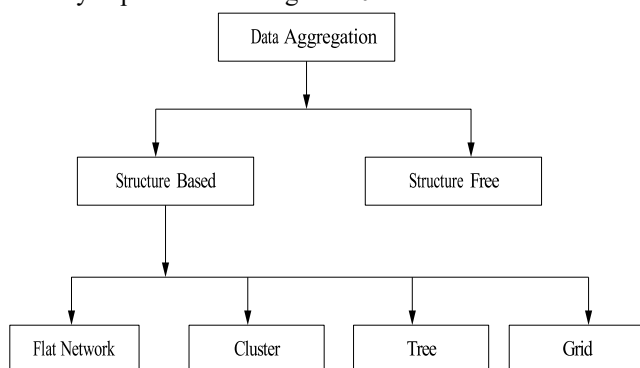


Figure 1.0 Taxonomy of Data Aggregation

II. Data Aggregation In Flat Networks

In flat networks, each sensor node 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 sensor nodes that have data matching the query send response messages back to the sink. The choice of a particular communication protocol depends on the specific application at hand. In the rest of this subsection the protocols used for data aggregation in flat network along with their advantages and limitations.

A. Flooding and gossiping

Flooding and gossiping are two classical mechanisms to relay data in sensor networks without the need for any routing algorithms and topology maintenance. In flooding, each sensor node receiving a data packet broadcasts it to all of its neighbors 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 neighbor that picks another random neighbor to forward the packet and so on.

b. Directed Diffusion

Directed diffusion (DD)[4] has been proposed by Heinzelman et al (1999) is a popular data aggregation paradigm for wireless sensor networks. It is a data-centric and application aware paradigm, in the sense that all the data generated by sensor nodes are named by attribute-value pairs.

Such a scheme combines the data coming from different sources en-route 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, which describes a required task to be implemented by the network. 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 neighbor 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.

B. Sensor Protocol For Information Via Negotiation (Spin)

SPIN [5] is among the early work to pursue data centric routing mechanism. The idea behind SPIN is to name the data using high-level descriptors or meta-data. Before transmission, metadata is exchanged among sensor nodes via a data advertisement mechanism, which is the key feature of SPIN. Each node upon receiving new data, advertises it to its neighbors and interested nodes. Here, neighboring nodes are those who do not have the data, retrieve the data by sending a request message. SPIN's meta-data negotiation solves the classic problems of flooding such as redundant information passing, overlapping of sensing areas and resource blindness thus, achieving a lot of energy efficiency. There is no standard meta-data format and it is assumed to be application specific, e.g. using an application level framing. There are three messages defined in SPIN to exchange data between nodes. They are: ADV message to allow a sensor to advertise a particular meta-data, REQ message to request the specific data and DATA message that carry the actual data.

d. Rumor Routing

Rumor routing is another variation of Directed Diffusion and is mainly intended for contexts in which geographic routing criteria are not applicable. Generally Directed Diffusion floods the query to the entire network when

there is no geographic criterion to diffuse tasks. However, in some cases there is only a little amount of data requested from the nodes and thus the use of flooding is unnecessary. An alternative approach is to flood the events if number of events is small and number of queries is large. Rumor routing is between event flooding and query flooding. The idea is to route the queries to the nodes that have observed a particular event rather than flooding the entire network to retrieve information about the occurring events. In order to flood events through the network, the rumor routing algorithm employs long-lived packets, called agents. When a node detects an event, it adds such event to its local table and generates an agent. Agents travel through the network in order to propagate information about local events to distant nodes. When a node generates a query for an event, the nodes that know the route, can respond to the query by referring its event table. Hence, the cost of flooding the whole network is avoided. Rumor routing maintains only one path between source and destination as opposed to Directed Diffusion where data can be sent through multiple paths at low rates.

e. Gradient-Based Routing

Gradient-Based Routing [10] is another version of directed diffusion that aims to distribute traffic evenly throughout the network to enhance the network lifetime. The key idea is to memorize the number of hops when the interest is diffused through the whole network. As such, each node can calculate a parameter called the height of the node, which is the minimum number of hops required to reach the base station. The difference between a node's height and that of its neighbor is considered the gradient on that link. A packet is forwarded on a link with the largest gradient.

III. Data Aggregation in Hierarchical Networks

As mentioned earlier about flat network may result in excessive communication and computation burden 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.

IV. Data Aggregation in Cluster-Based Network

In energy-constrained sensor networks of large size, it is inefficient for sensor nodes to transmit the data directly to the sink. In such scenarios, sensor nodes 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 in significant energy savings for the energy constrained sensors. Figure 1.1 shows a cluster based sensor network organization. The cluster heads can communicate with the sink directly via long range transmissions or multi-hopping through other cluster heads. Recently, several cluster-based network organization and data-aggregation protocols have been proposed. In this section to discuss various clustering protocols.

• LEACH

LEACH protocol is the first clustering protocol. It provides a conception of round. LEACH protocol runs with many rounds. Each round contains two states: cluster setup state and steady state. In cluster setup state, it forms cluster in self-adaptive mode; in steady state, it transfers data. The time of second state is usually longer than the time of first state for saving the protocol payload.

• E-LEACH

Energy-Leach which improves the CH selection procedure. Like LEACH protocol, E-LEACH protocol also divides into rounds, in the first round, every node has the same probability to turn into CH, that mean nodes are randomly selected as CHs, in the next round, the residual energy of each node is different after one round of communication and taken into account for the selection of the CHs. The nodes having more energy will become CHs rather than nodes with less energy.

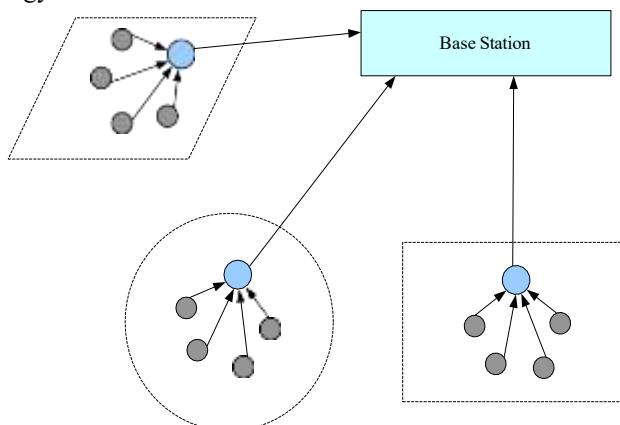


Figure 1.1 Cluster Based Data Aggregation

• TL-LEACH

In the LEACH protocol CH collects the information from cluster member nodes and after aggregation sends the information directly to the base station. CH might be located far away from the base station in that case it would be more energy consuming to send the information directly to the base station and CH will die quickly than other nodes. A new version of LEACH called Two-level Leach has been proposed in 2005[8]. In this protocol; CH collects information from cluster member and in spite of sending it directly to base station it sends it to another CH that lie between the CH and BS as a relay station.

• M-LEACH

In LEACH CH sends the aggregated information directly to the base station that is more energy consuming. In M-LEACH multi-hop communication is selected among CH. Then, according to the selected optimal path, these CHs transmit data to the corresponding CH which is nearest to BS. Finally, this CH sends data to BS. M-LEACH protocol is almost the same as LEACH protocol, only makes communication mode from single hop to multi-hop between CHs and BS.

• LEACH-C

LEACH-C[3] protocol which uses a centralized algorithm. LEACH-C protocol can produce better performance by dispersing the cluster heads throughout the network. During the set-up phase of LEACH-C, each node sends information about its current location (possibly determined using GPS) and residual energy level to the sink. In addition to determining good clusters, the sink needs to ensure that the energy load is evenly distributed among all the nodes. To do this, sink computes the average node energy, and determines which nodes have energy below this average. The steady-state phase of LEACH-C is identical to that of the LEACH protocol.

IV. Chain-Based Data Aggregation

In cluster-based sensor networks, sensor nodes transmit data to the cluster head where data aggregation is performed. However, if the cluster head is far away from the source sensor nodes, they might spend excessive energy in communication. Further improvements in energy efficiency can be obtained if sensor nodes transmit only to close neighbors. The key idea behind chain-based data aggregation is that each sensor node transmits only to its closest neighbor. A chain-based data-aggregation protocol called Power-Efficient Data Gathering Protocol for Sensor Information Systems (PEGASIS)[7]. In PEGASIS, nodes are organized

into a linear chain for data aggregation. The nodes can form a chain by employing a greedy algorithm or the sink can determine the chain in a centralized manner. Greedy chain formation assumes that all nodes have global knowledge of the network. The farthest node from the sink initiates chain formation and, at each step, the closest neighbor of a node is selected as its successor in the chain. In each data-gathering round, a node receives data from one of its neighbors, aggregates the data with its own, and transmits the aggregated data to its other neighbor along the chain. Eventually, the leader node which is similar to cluster head transmits the aggregated data to the sink. Figure 1.2 shows the chain-based data-aggregation procedure in PEGASIS. The greedy chain formation approach used may result in some nodes having relatively distant neighbors along the chain. This problem is alleviated by not allowing such nodes to become leaders.

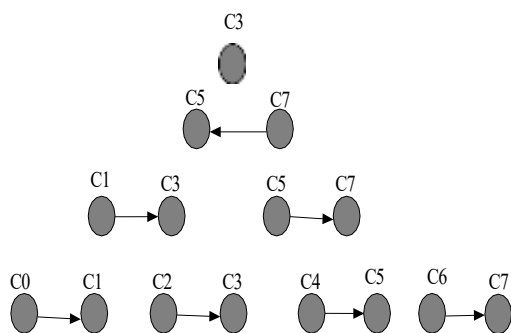


Figure 1.2 Chain based data aggregation

V. Tree-Based Data Aggregation

In a tree-based network, sensor nodes are organized into a tree where data aggregation is performed at intermediate nodes along the tree and a concise representation of the data is transmitted to the 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. Various tree data aggregation algorithms have been proposed in the literature. An Energy-Aware Data Aggregation Tree (EADAT) [9] where the base station (root) sends a broadcast control message periodically. Upon receiving this message for the first time, each node will start a timer. The expiration time is inversely proportional to the node's residual energy. The timer is refreshed when a node receives this message during the timer count down.

E-Span is an energy-aware spanning tree algorithm as shown in Figure 1.3. In E-span, the source node which has the

highest residual energy is chosen as the root. Other source nodes choose their corresponding parent node among their neighbors based on the information of the residual energy and distance to the root. If there are multiple neighbors with equal distance, the node which has most remaining energy is selected as parent. As E-Span protocol considers distance as main parameter and remaining energy as second, the network coverage is not high; because in some cases the nodes with low remaining energy are selected as parent. After local aggregation and data transmission, the remaining energy of these nodes is finished quickly. This causes the node failure and network cannot cover the region completely.

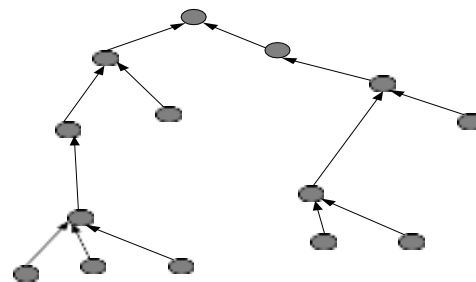


Figure 1.3 E-SPAN

VI. Grid-Based Data Aggregation

The data-aggregation schemes which are based on dividing the region monitored by a sensor network into several grids. They are: grid-based data aggregation[11] and In-network data aggregation. In grid-based data aggregation, a set of sensor nodes are assigned as data aggregators in fixed regions of the sensor network. The sensor nodes in a particular grid transmit the data directly to the data aggregator of that grid. Hence, the sensor nodes within a grid do not communicate with each other. In-network aggregation is similar to grid-based data aggregation with two major differences, namely, each sensor within a grid communicates with its neighboring sensors. Any sensor node within a grid can assume the role of a data aggregator in terms of rounds until the last node dies.

VII. Structure free data aggregation

In structure free data aggregation, it is not maintain any structure. This method is very useful in event based applications where event region changes very frequently and if structure based approach then to maintain the structure again and again. In structure free environment because it is not maintained any structure then to reconstruct the structure at the time of node failure or the changing of event region. There are two main challenges in performing structure free data aggregation. First, as there is no pre constructed structure, routing decisions for the efficient aggregation of packets need to be made on-the-fly. Second, as nodes do not explicitly

know their upstream nodes, they cannot explicitly wait on the data from any particular node before forwarding their own data.

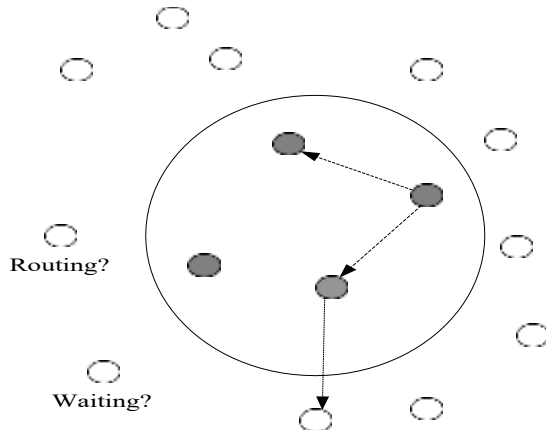


Figure 1.4 structure free data aggregation

The benefit of this approach is that it doesn't have to maintain the structure all the time whereas in structured environment to reconstruct the structure at the time when some nodes fail due to energy failure. The first work about the structure free data aggregation [6] where the authors observe that packets need to be aggregated early on their route to the sink for efficiency. Based on this observation, they propose and model a MAC layer protocol for spatial convergence called Data-Aware Anycast (DAA). Second, they observe that, if some nodes wait for other nodes to send data, it can lead to efficient aggregation. They work the impact of Randomized Waiting (RW) for improved data aggregation. In DAA a source node sends the Request To Send (RTS) packet to all of its neighbors, with RTS it also attaches the type of data it has sensed. After receiving the RTS only those neighbor nodes send CTS packet that have same type of data. After receiving the CTS from more than one neighbor, source node selects only one of them according to instantaneous channel condition. DAA is based on MAC layer any casting where the situation to select only one next hop among many. DAA improves the performance of data aggregation in comparison to structured approaches. If DAA with the RW is used, it further improves the performance.

VIII. Handling trade-offs in data aggregation

The performance of data-aggregation protocols are characterized by performance measures such as energy consumption, latency, and data accuracy. There is usually a trade-off between the different objectives. It section describes approaches for handling the trade-offs in data-aggregation schemes. Boulis et al [1] have studied the energy-accuracy trade-offs for periodic data-aggregation problems in sensor networks. They have considered a problem in which the sensors provide periodic estimates of the environment. A

distributed estimation algorithm has been developed which uses the "max" aggregation function. Some of the unique features of the proposed estimation algorithm are scalability with the network architecture and time.

The key idea of their approach is a threshold based scheme where the sensor nodes compare their fused estimates to a threshold to make a decision regarding transmission. However, in the absence of prior information about the physical environment, setting the threshold is a nontrivial task. The threshold can be used as a tuning parameter to characterize the tradeoff between accuracy and energy consumption. The main advantage of the proposed approach is that it does not depend on a hierarchical tree structure for performing data aggregation. Instead, every node has the global information about the aggregated data. The main disadvantage of the approach is that the functionality of the fusion algorithm depends on the aggregation function. Hence the fusion algorithm is not applicable for a wide range of aggregation functions such as "average," "count" or "min." Duarte-Melo&Liu [2] have studied the transport capacity of data gathering sensor networks with different communication organizations. The hierarchical and flat organizations of sensor networks were compared in terms of capacity and energy consumption. They have discussed the trade-offs between capacity and energy consumption for data-aggregation applications in which every sensor node sends an equal amount of original data to the sink. In the flat architecture, nodes communicate with the sink via multihop routes by using peer nodes as relays. In the hierarchical structure, nodes are organized into clusters where the cluster heads serve as simple relays for transmitting the data. For a hierarchical network, where cluster heads have the same transmission capacity as the sources, the minimum requirement on the number of clusters has been obtained for achieving the upper bound on the throughput. The main finding of their work is that higher throughput can be achieved by using clustering at the cost of the extra nodes functioning as cluster heads.

IX. Summary

A comprehensive discussion of data-aggregation techniques in wireless sensor networks. Here the taxonomy of data aggregation based on network topology is presented first. Then the comprehensive work of various data aggregation algorithms are presented. All of them focus on optimizing important performance measures such as network lifetime, data latency, data accuracy, and energy consumption. Efficient organization, routing, and data-aggregation tree construction are the three main focus areas of data aggregation techniques. The advantages and disadvantages of various data aggregation

techniques are described . Then it put emphasis on various trade-offs in data aggregation algorithms and at the end to review security issues in data aggregation in WSN. Most of the existing work has mainly focused on the development of an efficient routing mechanism for data aggregation. However, the performance of the data-aggregation technique is strongly coupled with the infrastructure of the network. There is no significant research in exploring the impact of heterogeneity, aggregator node selection and mode of communication (single hop versus multihop) and on the performance of the data-aggregation protocols. Although, many of the data-aggregation techniques discussed are promising, but still there is significant scope to work towards them by combining some of the aspects such as network lifetime, data latency and packet delivery ratio by exploring different approaches which will be discussed in the following chapters .

References

- 1 Boulis A, Ganeriwal S & Srivastava, MB 2003, 'Aggregation in sensor networks: an energy-accuracy trade-off, 1st IEEE Int'. Wksp. Sensor Network Protocols and Applications, USA
- 2 Duarte-Melo, EJ & Liu, M 2003, 'Data-gathering wireless sensor networks: organization and capacity Computer Networks: the Int', Computer and Telecommun. Net., pp.43,
- 3 Haslizan A, Halim1, Ab & Zen, K 2008, 'MAC Protocol to Reduce Packet Collision in Wireless Sensor Network', Proceedings of the International Conference on Computer and Communication Engineering.
- 4 Heinzelman, W, Kulik, J & Balakrishnan, H 1999,' Adaptive protocols for information dissemination in wireless sensor networks', in: The Proceedings of the 5th Annual ACM/ IEEE International Conference on Mobile Computing and Networking (MobiCom_99), Seattle, WA
- 5 Intanagonwiwat, C, Govindan, R & Estrin, D 2000, 'Directed diffusion: a scalable and robust communication paradigm 800 for sensor networks', iThe Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom_00), Boston, MA.
- 6 Kai -Wei Fan, Sha Liu & Prasun Sinha 2007, Structure-Free Data Aggregation in Sensor Networks, IEEE Transaction On Mobile Computing, vol.6, no. 8.
- 7 Lindsey, S & Raghavendra, CS 2002, 'PEGASIS: Power-efficient Gathering in Sensor Information System', Proceedings IEEE Aerospace Conference, vol. 3, pp. 1125-1130.
- 8 Loscri, V, Morabito, G & Marano, S 2005, 'A Two-Levels Hierarchy for Low-Energy Adaptive Clustering Hierarchy (TL-LEACH)', IEEE
- 9 Min Ding, Xiuzhen Cheng & Guoliang Xue 2006, 'Energy-Aware Data Aggregation Tree (EADAT) algorithm, Proceeding Mobility '06 Proceedings of the 3rd international conference on Mobile technology, applications & systems, ISBN:1-59593-519-3, doi> 10.1145/ 1292331.1292391.
- 10 Schurgers, C & Srivastava, MB 2001, 'Energy efficient routing in wireless sensor networks', The MILCOM Proceedings on Communications for Network-Centric Operations: Creating the Information Force, McLean, VA
- 11 Vaidhyanathan K 2004, 'Data Aggregation Techniques in Sensor Networks,' Technical Report, OSU-CISRC-11/04-TR60, Ohio State University