Approaches for Data Aggregation in Wireless Sensor Networks: A Survey

[1] **Shini K** , [2] **Shijukumar P S**

[1] M.Tech Student in CSE, Calicut University, Kerala, India

[2] Asst. Professor in CSE, Calicut University, Kerala, India

Abstract - Wireless Sensor Network (WSN) consists of many sensor nodes, which are capable of sensing data and send the sensed data to the base station. Sensor nodes are battery limited and most of the energy is consumed in transmission and reception of the data packets. Reduction in transmission of data packets greatly improves the lifetime of WSNs as it not only reduces the energy of transmitting nodes but also of the receiving nodes. Thus power efficient methods must be employed for data gathering and aggregation in order to achieve long time network usage. There are various approaches for data aggregation like LEACH, TEEN, PEGASIS, EADAT, PEDAP, PEDAP-PA, DASDR, L-PEDAP, EEDGP, and ADA.

***Keywords* - Wireless sensor networks, Data aggregation, LEACH, TEEN, PEGASIS, EADAT, PEDAP, PEDAP-PA, DASDR, L-PEDAP, EEDGP, ADA.**

1. Introduction

Wireless sensor network consists of small light-weighted wireless nodes called sensor nodes, deployed in physical or environmental condition. Sensor node measures physical parameters such as sound, pressure, temperature, and humidity. These nodes are deployed in large or thousand numbers and collaborate to form an ad-hoc network capable of reporting to data collection sink (base station).

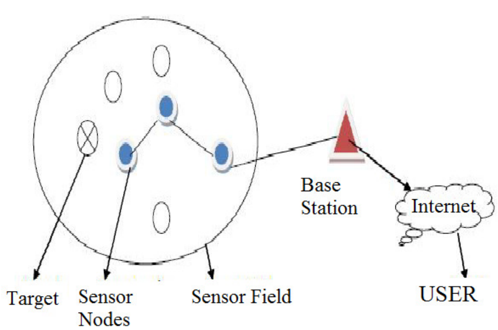

Fig. 1 Wireless Sensor Networks

Wireless sensor network have various applications like habitat monitoring, building monitoring, health monitoring, military surveillance and target tracking. However wireless sensor network is resource constraint if we talk about energy, computation, memory and limited communication capabilities. All sensor nodes in the wireless sensor network interact each other or with intermediate nodes. Figure 1 shows the architecture of wireless sensor networks.

Data Aggregation: In typical wireless sensor networks, sensor nodes are usually resource-constrained and battery-limited. In order to save resources and energy, data must be aggregated to avoid overwhelming amounts of traffic in the network. There has been extensive works on data aggregation schemes in sensor networks. The aim of data aggregation is to eliminate redundant data transmission and enhances the lifetime of wireless sensor networks. Figure 2 shows data aggregation and no data aggregation

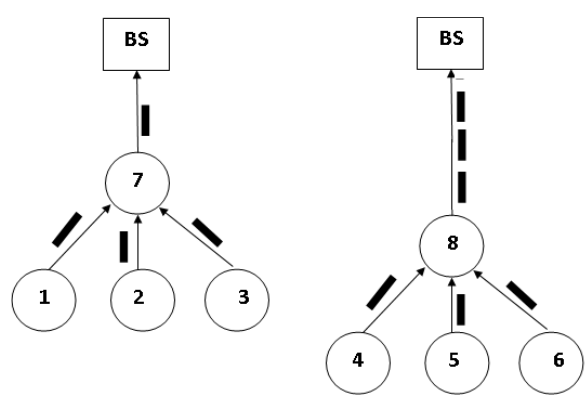

Fig. 2 Data Aggregation and No Data Aggregation

2. Literature Survey

Researchers are always being conducted to improve the accuracy and efficiency of data aggregation to improve network lifetime of wireless sensor networks. Some approaches to data aggregation in sensor networks are described here.

2.1 Low-Energy Adaptive Clustering Hierarchy (LEACH)

Wendi Rainer Heinzelman et al. [1] proposed a clustering-based protocol, LEACH (Low-Energy Adaptive Clustering Hierarchy), which is able to distribute energy dissipation evenly throughout the sensors, doubling the useful system lifetime for the networks. LEACH is a hierarchical based routing protocol which uses random rotation of the nodes required to be the cluster-heads to evenly distribute energy consumption in the network.

LEACH arranges the nodes in the network into small clusters and chooses one of them as the cluster-head. Node first senses its target and then sends the relevant information to its cluster-head. Then the cluster head aggregates and compresses the information received from all the nodes and sends it to the base station. Remaining nodes are cluster members. LEACH assumes that the nodes are homogeneous and the routing of packets to the base station is done in a single hop via the cluster-heads.

This protocol is divided into rounds; each round consists of two phases.

- Set-up phase
- Steady-state phase

Set-up Phase: In the set-up phase, a sensor node selects random number between 0 and 1. If this number is less than the threshold T(n), the node becomes a CH. T(n) is computed as:

$$T(n) = p/[1-p*(r \bmod 1/p)] \quad \text{if } n \in G$$
$$T(n) = 0 \quad \text{otherwise}$$

Where r is the current round; p, the desired percentage for becoming CH; and G is the collection of nodes not elected as a CH in the last 1/ p rounds.

Each node decides independent of other nodes if it will become a CH or not. This decision takes into account when the node served as a CH for the last time (the node that hasn't been a CH for long time is more likely to elect itself than nodes that have been a CH recently). In the following advertisement phase, the CHs inform their neighbourhood with an advertisement packet that they become CHs. Non-CH nodes pick the advertisement packet with the strongest received signal strength. In the next cluster setup phase, the member nodes inform the CH that they become a member to that cluster with "join packet" contains their Ids using CSMA. After the cluster-setup sub phase, the CH knows the number of member nodes and their IDs. Based on all messages received within the cluster, the CH creates a TDMA schedule, pick a CSMA code randomly, and broadcast the TDMA table to cluster members. After that steady-state phase begins.

Steady-State Phase: Data transmission begins; Nodes send their data during their allocated TDMA slot to the CH. This transmission uses a minimal amount of energy (chosen based on the received strength of the CH advertisement). The radio of each non-CH node can be turned off until the nodes allocated TDMA slot, thus minimizing energy dissipation in these nodes. When all the data has been received, the CH aggregate these data and send it to the BS.

Although LEACH protocol prolongs the network lifetime in contrast to plane multi-hop routing and static routing, it still has problems. The cluster heads are elected randomly, so the optimal number and distribution of cluster heads cannot be ensured. The nodes with low remaining energy have the same priority to be a cluster head as the node with high remaining energy. Therefore, those nodes with less remaining energy may be chosen as the cluster heads which will result that these nodes may die first. The cluster heads communicate with the base station in single-hop mode which makes LEACH cannot be used in large-scale wireless sensor networks for the limit effective communication range of the sensor nodes.

Applications of LEACH protocol are Fault detection and diagnosis.

2.2 Threshold sensitive Energy Efficient sensor Network Protocol (TEEN)

Arati Manjeshwar et al. [2] modified the work of Huseyin Ozgur et al.[1] and proposed a hierarchical protocol, Threshold sensitive Energy Efficient sensor Network protocol (TEEN) that uses data-centric mechanism. TEEN is well suited for time critical applications and is also quite efficient in terms of energy consumption and response time. In this nodes are arranged in hierarchical clustering scheme in which some nodes acts as 1st level and 2nd level cluster heads.

After forming the cluster head, it gets the attribute from the user. Once the attribute is received the cluster head broadcasts the attribute, hard threshold (HT) and soft threshold (ST) values to its cluster members. The sensor nodes starts sensing and transmits the sensed data when it exceeds HT. HT is the minimum attribute range above which the values are expected. The transmitted sense value is stored in an internal variable called sensed value (SV). The cluster nodes again starts sensing, when its value exceeds the ST i.e. the minimum change in the sensed value it switches on its transmitter and transmits. In TEEN the energy is conserved since the sensor nodes in the cluster senses continuously but transmits only when the sensed value is above HT. The ST further reduces the transmission which could have been occurred when there is a little change (or) no change in sensed attribute.

As the cluster heads need to perform extra computations it consumes more energy compared to other nodes. In order to evenly distribute the energy consumption each nodes in the cluster is given a chance to act as a cluster head for a fixed cluster period. The attributes can also be changed during every cluster change time.

The existing problems of TEEN is: on the one hand, if the node monitoring data has not been able to exceed the set hard threshold, the node will transmit the data, the user will not be able to get any data, also do not know whether the node failure; on the other hand, the node monitor to the appropriate data will be transmitted in real time data, using TDMA mechanism will cause the data delay.

2.3 Power Efficient Gathering in Sensor Information System (PEGASIS)

Stephanie Lindesy et al. [3] proposed a chain-based protocol, PEGASIS, which is able to increase the life-time of the network twice as much the lifetime of the network under the LEACH protocol.

PEGASIS protocol forms a chain of sensor nodes, where each sensor node only communicates with their neighbours. Sensor nodes are deployed in harsh physical environment. They have very limited computation capability because they are limited by the battery power. It has been a challenge to maximize the use of energy of these sensor nodes to extend the network lifetime.

The key idea in PEGASIS is to form a chain among the sensor nodes so that each node will receive from and transmit to a nearest neighbour. Gathered data move from node to node, get fused and eventually a designated node transmits to the base station (BS). Nodes take turns transmitting to the BS so that the average energy spent per round is reduced.

PEGASIS introduces excessive delay for distant node on the chain. In addition the single leader can become a bottleneck, which causes decreases of network lifetime WSNs. For example, every sensor needs to be aware of the status of its neighbour so that it knows where to route that data. Such topology adjustment can introduce significant overhead especially for highly utilized networks. This approach is also not very efficient, since the transmission distances can be quite long and finding a minimum distance chain is NP-complete (travelling salesman problem).

2.4 Energy Aware Distributed Aggregation Tree (EADAT)

Min Ding et al. [4] suggested an efficient energy- aware distributed heuristic to generate the aggregation tree, EADAT. EADAT performs very well in terms of network lifetime, energy saving, data delivery ratio and the protocol overhead. In this, sensors with higher residual power have a higher chance to become a non-leaf tree node and thus extend the network lifetime in terms of the number of live nodes.

The algorithm is initiated by the sink which broadcasts a control message. The sink assumes the role of the root node in the aggregation tree. The control message has five fields: ID, parent, power, status and hopcnt indicating the sensor ID, its parent, its residual power, the status (leaf, non-leaf node or undefined state) and the number of hops from the sink. After receiving the control message for the first time, a sensor v sets up its timer to Tv. Tv counts down when the channel is idle. During this process, the sensor v chooses the node with the higher residual power and shorter path to the sink as its parent. This information is known to node v through the control message. When the timer times out, the node v increases its hop count by one and broadcasts the control message. If a node u receives a message indicating that its parent node is node v, then u marks itself as a non leaf node. Otherwise the node marks itself as a leaf node. The process continues until each node broadcasts once and the result is an aggregation tree rooted at the sink.

The main advantage of this algorithm is that sensors with higher residual power have a higher chance to become a non-leaf tree node. To maintain the tree, a residual power threshold Pth is associated with each sensor. When the residual power of a sensor falls below Pth, it periodically broadcasts help messages for Td time units and shuts down its radio. A child node upon receiving a help message, switches to a new parent. Otherwise it enters into a danger state. If a danger node receives a hello message from a neighbouring node v with shorter distance to the sink, it invites v to join the tree.

EADAT prolongs network lifetime and saves more energy in comparison with routing methods without aggregation. The average energy level of sensor nodes decreases much more slowly compared to the scenario without data aggregation. The main disadvantages of EADAT are the extensive use of timers and require prior knowledge or support from a tree root.

2.5 Power Efficient Data gathering and Aggregation Protocol and its Power Aware version (PEDAP)

Huseyin Ozgur et al. [5] proposed Power Efficient Data gathering and Aggregation Protocol (PEDAP), which outperforms previous approaches, by constructing minimum energy consuming routing for each round of communication.

Power Efficient Data Gathering and Aggregation Protocol (PEDAP) is a Centralized routing protocol. Compared to LEACH and PEGASIS, PEDAP achieves improvement in network lifetime. The defined constraint of sensor nodes is their very low finite battery energy, which limits the lifetime and the quality of the network. For that reason, the protocols running on sensor networks must consume the resources of the nodes efficiently in order to achieve a longer network lifetime.

In PEDAP the packet routes is based on Minimum cost spanning Tree (MST), which improves the lifetime of the system and node energy is directly related to it degree and the distance to its parents. In single round PEDAP takes minimum amount of energy. Additionally the authors propose a new version of power-aware called PEDAP-PA. It computes remaining energy of the sender using cost function.

$$c_{ij(k)} = \frac{2 * E_{elec} * k + E_{amp} * k * d_{ij}^2}{e_i} \qquad (1)$$

$$c_i^{\prime}(k) = \frac{E_{elec} * k + E_{amp} * k * d_{ib}^2}{e_i} \qquad (2)$$

Where Cij is the transmission between node i and j. C′i is the cost between node i and the sink. Dij is the cost transmission between node i and j. dib is the distance between node i and the sink. The energy dissipation of the radio in order to run the transmitter or receiver circuitry is equal to Eelec =50nJ\bit, and to run the transmit amplifier it is equal to Eamp=100pJ\bit\m2. Since C′i is smaller than Cij when the term with Eamp is much smaller than the term with Eelec. From Equation (1) and (2) the advantage is overall lifetime and it increases the number of transmissions to the sink.

PEDAP extends the lifetime of the last node by minimizing the total data gathering in each round where as PEDAP-PA balances the energy consumption among the nodes. Edge cost is computed as sum of transmission and receiving energy in PEDAP. In PEDAP-PA considering the cost by dividing PEDAP edge cost with transmitter residual energy. The disadvantages of PEDAP and PEDAP-PA are however centralized in nature. This scheme is focusing only Shortest Path. It fails to achieve Bandwidth Utilization.ie, it unable to improve the Resource Utilization load; it recomputed the routing tree after a predefined number of rounds. In PEDAP and PEDAP-PA Edge weight assignment is calculated with only transmitters' residual energy. This is the major drawback to improve the reliability of the system.

2.6 Data Aggregation Supported by Dynamic Routing (DASDR)

Jiao Zhang et al. [6] proposed an effective data aggregation mechanism supported by dynamic routing (DASDR) which can adapt to different scenarios without incurring much overhead. DASDR is more effective in energy savings as well as scales well regarded to the network size. To overcome the drawbacks of the static routing, here proposed an efficient data aggregation mechanism supported by dynamic routing. The challenge of this method is how to make full use of the local information to make the routing decisions rapidly and ensure the decisions are efficiently adapt to both data gathering and event-based applications. This is achieved by introducing the concept of potential in classical physical theory into WSN. The contributions are described as follows.

- The potential based dynamic routing scheme in DASDR only needs local information to make routing decisions. It is low-cost, can rapidly make forwarding decisions and performs well in both static and mobile scenarios.
- In the dynamic routing scheme, a queue potential field is constructed by exploiting the local

information of queue length, which can make packets more spatially convergent.

DASDR can adapt to different scenarios without incurring much overhead. It is more effective in energy savings as well as scales well regarded to the network size.

2.7 Localized Power-Efficient Data Aggregation Protocol (L-PEDAP)

Huseyin Ozgur Tan et al. [7] modified their work, in [5] as a localized version of PEDAP (L-PEDAP), which tries to combine the desired features of MST and shortest weighted path-based gathering algorithms. The main concern in this work is the lifetime of the network.

The routing approach in L-PEDAP consists of two phases. In the first phase, it computes a sparse topology over the visibility graph of the network in a localized manner. In the second phase, it computes a data gathering tree over the edges of the computed sparse topology. The topology needs to be efficiently computed by using only the one-hop neighbourhood information. For the first phase, two different sparse topologies are tested in a distributed manner, namely, local minimum spanning tree (LMST) and relative neighbourhood graph (RNG). These structures are supersets of MST and can be efficiently computed in a localized manner. For the second phase, three different methods are proposed. All of the methods are based on flooding a special packet using only the edges of the computed structure. According to the decisions made during this flooding process, the tree is yielded. These three methods that can be executed at a node for choosing the parent node toward the sink are to choose: 1) the first node from which the special packet is received, 2) the node that minimizes the number of hops to the sink, and 3) the node that minimizes the total energy consumed over the path to the sink.

L-PEDAP is adaptive since it considers the dynamic changes when constructing a routing tree. This protocol considers remaining life-time of nodes to improve network life-time. Disadvantages of L_PEDAP are it does not have any mechanism to find the cost of setup and communication time is more.

2.8 Energy Efficient Data Gathering Protocol (EEDGP)

The method suggested by Siddhartha Chauhan et al. [8] describes an energy efficient data gathering protocol (EEDGP) which reduces the transmission of data packets thereby reducing the energy consumption of SNs (SNs). EEDGP significantly increases the life time of WSNs by reducing the transmission of redundant data.

The protocol energy efficient data gathering protocol (EEDGP) has been designed to improve the network lifetime by conserving the energy of the nodes. Here the focus is on data gathering application, where SNs report sampled data to sink through CHs. EEDGP reduces the transmission of new data by nodes if it is similar to the previous data. EEDGP also uses overhearing within the cluster in order to reduce the transmission thus conserving transmission energy of SNs and reception energy of their respective CHs.

EEDGP is based on a time division multiple access (TDMA) where MAC frame is divided into periodic time slots and each SNs is allocated a particular slot for transmission. Each SN can transmit one packet in its allocated time slot so that collisions are avoided. EEDGP can be implemented on any cluster based protocols for WSNs. It is quiet accurate and can be used for applications where some error is acceptable.

2.9 Attribute-aware Data Aggregation Scheme (ADA)

Fengyuan Ren et al. [9] introduced the concept of packet attribute, defined as the identifier of the data sampled by different kinds of sensors or applications, and then proposed an attribute-aware data aggregation (ADA) scheme consisting of a packet-driven timing algorithm and inspired by the concept of potential in physics and pheromone in ant colony, a potential-based dynamic routing is elaborated to support the ADA strategy. The ADA scheme can make the packets with the same attribute spatially convergent as much as possible and therefore improve the efficiency of data aggregation. Furthermore, the ADA scheme also offers other properties, such as scalable with respect to network size and adaptable for tracking mobile events.

The main contributions in this work are threefold:

- An ADA scheme is proposed to intentionally drive the packets with the same attribute convergent as much as possible in the WSNs with heterogeneous sensors or various applications.
- Inspired by the concepts of both potential field in physics and pheromone in ant colony, a dynamic routing protocol is elaborately designed to support the ADA scheme.

- An adaptive packet-driven timing control algorithm is proposed to provide more chances for data aggregation on nodes.

The ADA scheme works in an environment with heterogeneous sensors and improves the efficiency of data aggregation and hence the network lifetime.

3. Performance Analysis

Various methods for data aggregation are discussed here. Methods like LEACH, PEGASIS, TEEN, EADAT, PEDAP, PEDAP-PA, L-PEDAP, EEDGP uses homogenous sensors and static routing methods. DASDR also uses homogeneous sensors but uses dynamic routing method. ADA works in an heterogeneous environment and uses dynamic routing technique. A comparison of above methods is given in the following table.

Table 1: Comparison of Different Methods

Method	*Structure*	*Routing*	*Node Type*
LEACH	Cluster	Static	Homogeneous
TEEN	Cluster	Static	Homogeneous
PEGASIS	Chain	Static	Homogeneous
EADAT	Tree	Static	Homogeneous
PEDAP	Tree	Static	Homogeneous
PEDAP-PA	Tree	Static	Homogeneous
DASDR	Tree	Dynamic	Homogeneous
L-PEDAP	Tree	Static	Homogeneous
EEDGP	Cluster	Static	Homogeneous
ADA	Tree	Dynamic	Heterogeneous

4. Conclusions

Data aggregation is one of the main methods to conserve energy in wireless sensor network (WSN). Even though methods like LEACH, TEEN, EADAT etc, with static routing in homogeneous sensor networks can improve the efficiency of data aggregation and hence network life time, methods like DASDR and ADA uses dynamic routing and performs more better. Along with the concept of dynamic routing, ADA scheme aggregate data from heterogeneous sensors and uses the concept of packet attribute.

It can make the packets with the same attribute convergent as much as possible to further improve the efficiency of data aggregation and the network lifetime.

References

[1] A. Chandrakasan W. Heinzelman and H. Balakrishnan. "Energy-Efficient Communication Protocols for Wireless Micro sensor Networks". In Proceedings of Hawaiian International Conference on Systems Science, January 2000.

[2] A. Manjeshwar and D. P. Agrawal. "TEEN: A Protocol for Enhanced Efficiency in Wireless Sensor Networks". In the Proceedings of the 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, San Francisco, CA, April 2001.

[3] S. Lindsey, C.S. Raghavendra. "PEGASIS: Power-efficient Gathering in Sensor Information System". Proceedings IEEE Aerospace Conference, vol. 3,Big Sky,MT,Mar. 2002,pp.1125-1130.

[4] X. Cheng M. Ding and G. Xue. "Aggregation Tree Construction in Sensor Networks". IEEE Vehic. Tech. Conf., vol. 4, pp. 21682172, Oct.2003.

[5] H. O. Tan and I. Korpeoglu. "Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks". SIGMOD Record, vol.32, no. 4, Dec. 2003, pp. 6671.

[6] Fengyuan Ren Tao He Jiao Zhang, Qian Wu and Chuang Lin. "Effective Data Aggregation Supported by Dynamic Routing in Wireless Sensor Networks". In Proceedings of the IEEE International Conference on Communications (ICC10), pages 16, May 2010.

[7] Huseyin Ozgur Tan and Ivan Stojmenovic. "Computing Localized Power-Efficient Data Aggregation Trees for Sensor Networks". IEEE Transactions on Parallel and Distributed Systems, VOL. 22, NO. 3 March 2011.

[8] Dr. Lalit Kumar Awasthi Siddhartha Chauhan. "Energy Efficient Data Gathering Protocol for Wireless Sensor Networks". IJCST Vol. 2, Issue 4, Oct -Dec. 2011.

[9] Yongwei Wu Tao He Canfeng Chen Fengyuan Ren, Jiao Zhang and Chuang Lin. "Attribute-Aware Data Aggregation Using Potential-Based Dynamic Routing in Wireless Sensor Networks". IEEE Transactions on Parallel And Distributed Systems, VOL. 24, NO. 5, May 2013.

[10] Y. Sankarasubramaniam E. Cayirci I.F. Akyildiz, W. Su. "Wireless sensor networks: a survey". Computer Networks 38 (2002) 393422.

[11] Dr.M.Rajani M.Shankar, Dr.M.Sridar. "Performance Evaluation of LEACH Protocol in Wireless Network". International Journal of Scientific Engineering Research, Volume 3, Issue 1, January-2012 1 ISSN 2229-5518.

[12] Krishna M. Sivalingam Stephanie Lindsey, Cauligi Raghavendra. "Data Gathering Algorithms in Sensor Networks Using Energy Metrics". IEEE Transactions on Parallel AND Distributed Systems, VOL. 13, NO. 9, September 2002.

[13] X. Cao F. Hu and C. May. "Optimized Scheduling for Data Aggregation in Wireless Sensor Networks". Proc.

Intl Conf. Information Technology: Coding and Computing (ITCC 05), pp. 156- 168, 2005.

Shini K received the bachelor's degree in Information Technology from the Mahatma Gandhi University, Kerala in 2005. Presently she is pursuing her M.Tech in the department of Computer Science and Engineering from Calicut University, Kerala. She has teaching experience of more than six years. Her research interests include computer networks, wireless sensor networks etc.

Shijukumar P S received the master's degree in Computer Science and Information Technology from Manonmaniam Sundaranar University, Tamil Nadu in 2012. Currently working as a Professor in Computer Science and Engineering Department, MES College of Engineering, Under the Calicut University Kerala. Research intrests includes High speed Network, wireless Sensor Network and Language Processing etc.