ISSN: 2321-8169
838 – 845

Comparative Study On Data Aggregation Techniques for Wireless Sensor Networks

Abhijith H V
Dept. of Information Science and Engineering
BMS College of Engineering
Bangalore, India
abhijithhv.mithra@gmail.com

Dr M Dakshayini
Dept. of Information Science and Engineering
BMS College of Engineering
Bangalore, India
dakshayini.ise@bmsce.ac.in

Abstract— The Wireless Sensor Networks (WSN) is one of the emerging technologies in the field of wireless ad-hoc networks. It consists of several low cost and low power sensor nodes which are capable of sensing, processing and communicating the various environmental parameters. These sensor nodes are randomly and densely deployed in the region of interest. The denser deployment of sensor nodes leads to the sensing and transmission of redundant information. Routing of such redundant data not only saturates the network resources, but also results in the wastage of energy and hence reduces the network lifetime. Data aggregation is the techniques which aggregate the data from different sensor nodes and reduces the redundant transmissions. Data aggregation ensures the efficient utilization of energy and hence enhances the network lifetime. In this paper, we present a survey on different data aggregation techniques for Wireless Sensor Networks.

Keywords*- Data Aggregation; WSN; Network Lifetime; Energy; Sensor Node; Redundant Data.*

*****

I. INTRODUCTION

The Wireless Sensor Networks is one among the emerging networking technologies of 21st century. Wireless Sensor Networks have received attention from both academics and industry because of its wider application range. Wireless Sensor Network consists of a large number of low-power, and multifunctional sensor nodes, with sensing, wireless communications and computation capabilities. These sensor nodes can communicate over short distance via wireless medium and collaborate to accomplish a common task, for example, environment monitoring, military surveillance, and industrial process control.

The sensor nodes are energy constrained, therefore it is inefficient for all the sensor nodes to transmit the sensed data directly to the sink node. Data sensed by the sensor nodes which are nearer to each other is redundant. In addition, it is difficult for the sink node to process huge amount of data. Hence, there is a need for a method which combines the data from different sensor nodes and reduces the number of packets to be transmitted to the sink node. This results in the saving of energy and increase in the network lifetime. Wireless sensor nodes require less power for processing the data than compared to transmitting data. It is preferable to do in network processing inside network and reduce packet size. Since sensor nodes may generate significant redundant data, similar packets from multiple nodes can be aggregated so that the number of transmissions is reduced. This can be accomplished by data aggregation Techniques.

Data aggregation is a process of combining the data from multiple sensor nodes to avoid redundant transmission and provide aggregated information to the sink node. The Data aggregation attempts to collect the critical data from the neighboring and intermediate sensor nodes and make it available to the base station in an energy efficient manner with minimum data latency. Data latency is important in many applications such as environment monitoring, where the freshness of data is also an important factor.

The main goal of data-aggregation algorithms is to collect and aggregate data in an energy efficient manner so that network lifetime is increased.

The working of Data aggregation algorithm is shown in the figure 1. The data from different sensor nodes are passed to the data aggregation algorithm. The data aggregation algorithm aggregates the data based on the application by using different data aggregation functions such as, max, min, count, average, sum, concat. Then the aggregated data is transmitted to the sink node [1].

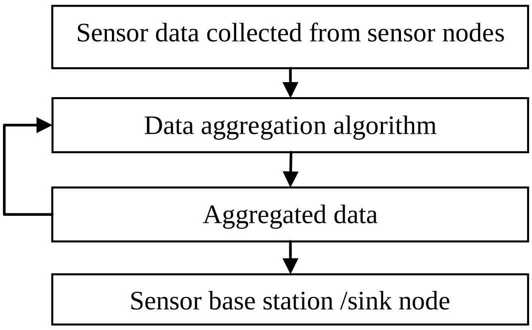

Figure 1: General architecture of the data aggregation algorithm

The remaining sections of the paper are structured as follows: Section II provides the performance characteristics of data aggregation techniques. Section III describes the classification of data aggregation techniques. Section IV describes the different data aggregation protocols. Section V draws the conclusions.

II. PERFORMANCE CHARACTERISTICS

A. 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. In many cases there may be chance of compromising data

ISSN: 2321-8169
838 – 845

accuracy. Effective algorithms with algebraic or statistical operations should be incorporated for maintaining data accuracy [2].

B. 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 [2].

C. Energy Efficiecy

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 collection round. A data aggregation scheme is energy efficient if it maximizes the functionality of the network. If we assume that all sensors are equally important, we should minimize the energy consumption of each sensor In addition, energy efficiency and network lifetime are synonymous in that improving energy efficiency enhances the lifetime of the network. Network lifetime quantifies the energy efficiency of the network [2].

D. 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].

III. CLASSIFICATION OF DATA AGGREGATION TECHNIQUES

The Data Aggregation in Wireless Sensor Networks can be classified into five basic types.

1) Centralized Data Aggregation
2) Cluster Based Data Aggregation
3) Multipath Data Aggregation
4) Tree Based Data Aggregation
5) In Network Data Aggregation

A. Centralized Data Aggregation

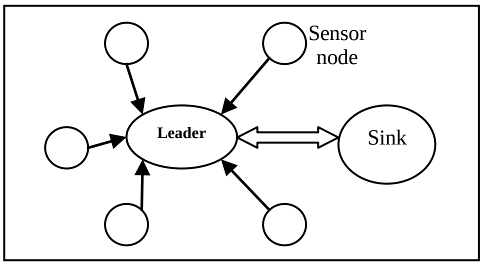

Figure 2: Centralized Data Aggregation

Centralized Data Aggregation is an address centric approach where each node sends data to a central node via the shortest possible route using a multi hop wireless protocol. The sensor nodes simply send the data packets to a leader, which is a powerful node. The leader aggregates the data which can be queried.

B. Cluster Based Data Aggregation

In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to sink.

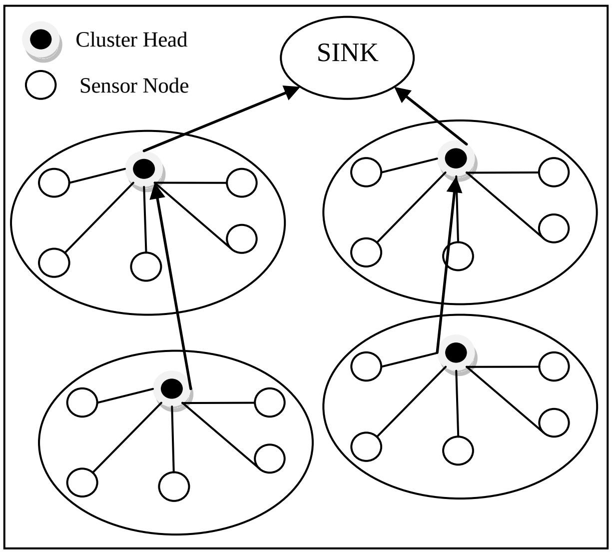

Figure 3: cluster based data aggregation

C. Multipath Data Aggregation

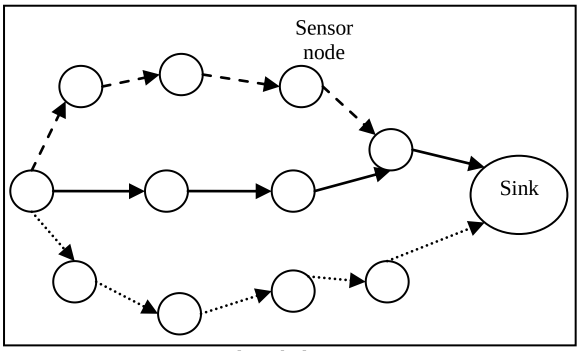

Figure 4: Multipath data aggregation

In Multipath Data Aggregation every nodes can send the data to its neighbor nodes. Data aggregation is performed in every intermediate node between source and sink. If the link or node fails Multipath data aggregation approach will discover alternative paths.

D. Tree Based Data Aggregation

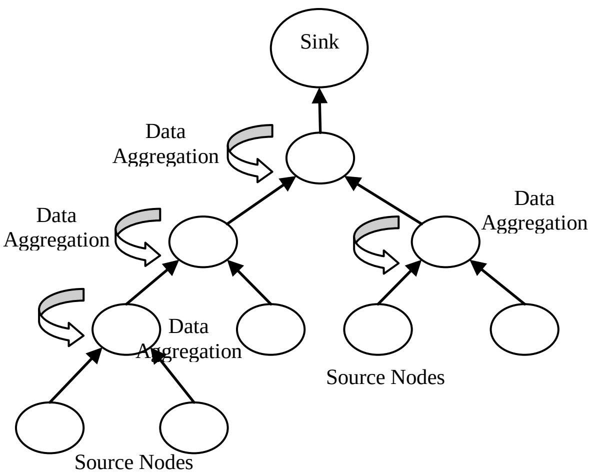

Figure 5: Tree Based data aggregation

ISSN: 2321-8169
838 – 845

In the tree-based approach aggregation is performed by constructing an aggregation tree, which could be a minimum spanning tree, rooted at sink and source nodes are considered as leaves. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and therein the aggregation done by parent nodes.

E. In-Network Data Aggregation

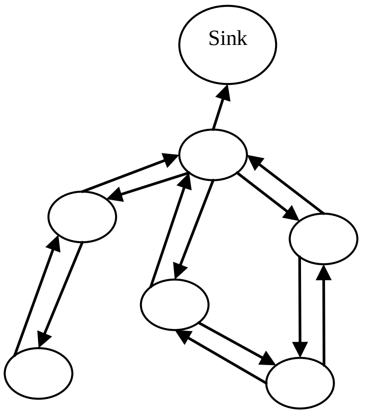

Figure 6: In-Network data aggregation

In-network aggregation is the global process of gathering and routing information through a multi-hop network, processed data at intermediate nodes with the objective of reducing resource consumption, thereby increasing network lifetime.

There are two approaches for in-network aggregation with size reduction and without size reduction. In-network aggregation with size reduction refers to the process of combining & compressing the data packets received by a node from its neighbors in order to reduce the packet length to be transmitted or forwarded towards sink. In-network aggregation without size reduction refers to the process merging data packets received from different neighbors in to a single data packet but without processing the value of data.

IV. Data Aggregation Protocols

The brief overview of different data aggregation protocols for Wireless Sensor Networks is discussed in this section.

A. Flat Networks based Data Aggregation protocols

In flat network, aggregation is performed in data centric routing method, the sink node transmits a query message to the other sensor nodes within the network, and sensor nodes which have data equivalent to the query send response messages back to the sink node. In this method excessive communications and computations are performed at the sink node hence this leads to faster depletion of its battery power. The failure of the sink node leads to failure of entire network.

There are two types of flat network protocols they are push diffusion and pull diffusion.

1) Push Diffusion:

In push diffusion the sensor nodes will transmit the sensed data towards sink node. The SPIN protocol is the example for push diffusion.

Sensor Protocol for Information via Negotiation (SPIN):

J. Kulik, W.R. Heinzelman and H. Balakrishnan proposed SPIN protocol [4]. The two main features of SPIN are negotiation and resource_adaptation. For successful data negotiation, sensor nodes need a descriptor to succinctly describe their observed data. These descriptors are defined as *metadata*. The format of the metadata is application specific.

SPIN nodes use three types of messages to communicate: ADV – new data advertisement. When a SPIN node has data to share, it can advertise this fact by transmitting an ADV message containing meta-data.REQ – request for data. A SPIN node sends an REQ message when it wishes to receive some actual data. DATA – data message. DATA messages contain actual sensor data with a meta-data header. The initiating node which has new data advertises the data to the neighboring nodes in the network using the metadata ADV. A neighboring node which is interested in this kind of data sends a request REQ to the initiator node for data. The initiator node responds and sends data to the sinks using DATA. Each node has a resource manager which keeps track of its energy usage. Each node polls its resources such as battery power before data transmission.

Advantage of SPIN is that topological changes are localized, since each node only requires the knowledge of its single-hop neighbors. Disadvantage of SPIN is its inability to guarantee data delivery.

2) Pull Diffusion:

In pull diffusion the sink node will send the query message to other nodes. Node which matches the query will reply back to sink node. There are two types of pull diffusion: two-phase pull diffusion [5] and one-phase pull diffusion [8].

Two-phase pull Diffusion:

C. Intanagonwiwat, R. Govindan and D. Estrin proposed the ***directed diffusion*** [5] protocol. The attributes of the data are utilized as message in the network. If the attributes of the data generated by the source matches the interest, a gradient is set to identify the data generated by the sensor nodes. The sink initially broadcasts an interest message in the network. The gradient specifies the data rate and the direction to send the data. Intermediate nodes are capable of caching and transforming the data. Each node maintains a data cache which keeps track of recently seen data items. After receiving low-data-rate events, the sink reinforces one particular neighbor in order to attract higher-quality data. Thus, directed diffusion is achieved by using data-driven local rules. Advantage of directed diffusion is, it is not necessary to maintain global network topology, unlike in SPIN. Disadvantage of Directed diffusion is, it is not suitable for applications which require continuous data delivery to the sink. Two -phase pull diffusion results in large overhead if there are many sources and sinks.

Rumor routing:

Braginsky D, Estrin D proposed Rumor routing [25] algorithm for sensor networks. Rumor Routing is variant of Directed Diffusion and it is used where geographic routing principles are not applicable. The idea is to route the queries to the nodes that have detected a particular event rather than

ISSN: 2321-8169
838 – 845

flooding the whole network to retrieve information about the going on events. In order to flood the events through the network, this routing algorithm uses packets called agents. When an event occurs, it reports such event to its local table and generates an agent. Agents travel the network in order to spread information about that event to distant nodes. Hence, the cost of flooding the whole network is to be avoided. Rumor routing maintains only one path between source and destination.

One-Phase Pull Diffusion:

B. Krishnamachari and J. Heidemann proposed One-phase pull-diffusion [8], which skips the flooding process of directed diffusion. In one-phase pull diffusion, sinks send interest messages that propagate through the network, establishing gradients. However, the sources do not transmit exploratory data. The sources transmit data only to the lowest latency gradient pertinent to each sink. Hence, the reverse route (From the source to the sink) has the least latency. The simulation results show that one-phase pull outperforms push diffusion when the source event rate is very high. When the sink interest rate is high, push diffusion performs better than one-phase pull diffusion.

B. Hierarchical Network Based Data Aggregation Protocols

In Hierarchical data aggregation protocols data aggregation is performed at special nodes, which reduces the number of messages transmitted to the sink. This improves the energy efficiency of the network. There are four types of hierarchical data aggregation they are the cluster based data aggregation, the grid based data aggregation, the chain based data aggregation and the tree based data aggregation.

1) Cluster Based Data Aggregation Protocols:

Low Energy Adaptive Clustering Hierarchy (LEACH):

LEACH [10] was proposed by W. R. Heinzelman 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. 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.

Hybrid Energy-Efficient Distributed clustering approach (HEED):

O. Younis and S. Fahmy proposed HEED protocol [11], with the main goal of forming efficient clusters for maximizing network lifetime. The main assumption in HEED is the availability of multiple power levels at sensor nodes. Cluster-head selection is based on a combination of node residual energy of each node and a secondary parameter which depends on the node proximity to its neighbors or node degree. The cost of a cluster head is defined as its average minimum reach ability power (AMRP). AMRP is the average of the minimum power levels required by all nodes within the cluster range to reach the cluster head. AMRP provides an estimate of the communication cost.

Energy Efficient Clustering and Data Aggregation (EECDA):

D. Kumar, T.C. Aseri, R.B. Patel proposed EECDA [13], which combines energy efficient cluster based routing and data aggregation for improving the performance in terms of lifetime and stability.

Cluster head election phase: EECDA considers three types of nodes (i.e., normal, advanced and super) which have deployed in a harsh wireless environment where battery replacement is impossible. Nodes with higher battery energy are advanced and super nodes and the remaining nodes are normal nodes. Intuitively, advanced and super nodes have to become CHs more often than the normal nodes.

Route selection: Once all CHs are elected in a specific round by using weighted election probability, each CH to estimate its energy residue and broadcast this information with its CH role to the neighboring nodes.

Data communication: In data communication phase, each non-CH node transmits its data to the associated CH. Each CH will receive all sensed data from its associated non-CH nodes and sends it to the BS. Therefore, each CH first aggregates the received data and then transmits the aggregated data to the Base Station (BS).EECDA has better network lifetime, stability and energy efficiency when compared with LEACH protocols.

Threshold sensitive Energy Efficient sensor Network (TEEN):

Manjeshwar, E.; Agrawal, D.P. proposed the Threshold sensitive Energy Efficient sensor Network [26]. TEEN is a cluster based hierarchical approach which follows the same procedure of LEACH protocol identify the cluster head. TEEN differs from LEACH by utilizing two different cluster heads: First level cluster head and second level cluster head. First level cluster heads are far away from base station and second level cluster heads are closer to the base station. The sensor nodes sense the information and send the sensed data to First level cluster head. First level cluster heads aggregate the data and transmit the aggregated to second level cluster head and finally to base station. It reduces the energy wastage of cluster head to reach the base station.

ISSN: 2321-8169
838 – 845

Energy Efficient Clustering Scheme (EECS):

M. Ye, C. Li, G. Chen, and J. Wu proposed Energy Efficient Clustering Scheme [27]. EECS is a clustering algorithm in which cluster head candidates contest for becoming cluster head by broadcasting their residual energy to neighbors. If the node does not find any other node with more residual energy, then it becomes a cluster head. Cluster formation is different than that of LEACH. EECS extends this algorithm by dynamic sizing of clusters based on cluster distance from the base station. It addresses the problem that clusters at a greater range from the base station requires more energy for transmission than those that are closer. Ultimately, this improves the distribution of energy throughout the network, resulting in better resource usage and enhance network lifetime.

2) Chain Based Data Aggregation Protocols:

In cluster-based sensor networks, sensors transmit data to the cluster head where data aggregation is performed. However, if the cluster head is far away from the sensors, they might expend excessive energy in communication. Further improvements in energy efficiency can be obtained if sensors transmit only to close neighbors. The key idea behind chain-based data aggregation is that each sensor transmits only to its closest neighbor.

Power-Efficient Data-Gathering Protocol for Sensor Information Systems (PEGASIS).

In PEGASIS [14] proposed by Lindsey S Raghavendra C, 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, fuses the data with its own, and transmits the fused 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. Advantage of PEGASIS protocol has considerable energy savings compared to LEACH. The main disadvantage of PEGASIS is the necessity of global knowledge of all node positions to pick suitable neighbors and minimize the maximum neighbor distance.

Chain-Based Hierarchical Routing Protocol (CHIRON):

Energy-efficient hierarchical chain-based routing protocol, named as CHIRON [16] was proposed by Kuong-Ho Chen , Jyh-Ming Huang, Chieh- Chuan Hsiao. 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. The operation of CHIRON protocol consists of four phases:

Group Construction Phase: The main purpose of this phase is to divide the sensing field into a number of smaller areas so that the CHIRON can create multiple shorter chains to reduce the data propagation delay and redundant transmission. It adopts the technique of Beam Star to organize its groups. After the sensor nodes are scattered, the BS gradually sweeps the whole sensing area, by successively changing different transmission power levels and antenna directions, to send control information to all nodes. After all nodes receiving such control packets, they can easily determine respective group.

Chain Formation Phase: In this phase, the nodes within each group will be linked together to form a chain respectively. For each group the node that is farthest away from the BS is initiated to create the group chain. By using a greedy algorithm, the nearest node will be chosen. The process is repeated until all nodes are put together, and thus finally a group chain is formed.

Leader Node Election Phase: For data transmission, a leader node in each group chain must be selected for collecting and forwarding the aggregated data to the BS. CHIRON chooses the chain leader based on the maximum value residual energy of group nodes. Initially, in each group, the node farthest away from the BS is assigned to be the group chain leader. After that, for each data transmission round, the node with the maximum residual energy will be elected. The residual power information of each node can be piggybacked with the fused data to the chain leader along the chain, so that the chain leader can determine which node is to be the new leader for next transmission round.

Data Collection and Transmission Phase: In the data collection and transmission phase the normal nodes in each group transmit their collected data through nearest nodes to the chain leader. And then, starting from the farthest groups, the chain leaders collaboratively relay their aggregated sensing information to the BS, in a multi-hop, leader-by-leader transmission manner.

Chain Oriented Sensor Network (COSEN):

N. Tabassum, Q. E. K. M. Mamun, and Q. Urano proposed the Chain Oriented Sensor Network [28] is a two-tier hierarchical chain-based routing scheme which operates in two phases. *Chain formation phase*: In this phase chains of different levels are formed. Sensor nodes are geographically grouped into several low-level chains. 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.

Data transmission phase: In this phase the information is transmitted along with the designated paths. Lower level leader nodes are responsible to collect information from lower level chains and send the information towards higher level leader. Higher level leader sends the information to BS. The sensor nodes are capable of dynamic power adjustment. Therefore nodes can adjust the amplifier electronics to adjust/accommodate for any required distance.

3) Tree Based Data Aggregation Technique:

Energy-aware distributed heuristic to construct and maintain a data-aggregation tree (EADAT):

M. Ding, X. Cheng and G. Xue proposed EADAT [17]. 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 sets up its timer. During this process, the sensor chooses the node with the higher residual power and shorter path to the sink as its parent. When the timer times out, the node increases its hop count by one and broadcasts the control message. If a node *x* receives a message indicating that its parent node is node *y*, then *x* 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.

Power-Efficient Data gathering and Aggregation Protocol (PEDAP):

The goal of PEDAP [18] which is proposed by H. O. Tan and I. Korpeoglu, is to maximize the lifetime of the network in terms of number of rounds, where each round corresponds to aggregation of data transmitted from different sensor nodes to the sink. PEDAP is a minimum spanning tree-based protocol which improves the lifetime of the network even when the sink is inside the field. PEDAP minimizes the total energy expended in each communication round by computing a minimum spanning tree over the sensor network with link costs. The main advantage of this algorithm is that sensors with higher residual power have a higher chance to become a non leaf tree node. Disadvantage of The PEDAP protocol requires global knowledge of the location of all nodes at the sink.

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

S.S. Satapathy and N. Sarma proposed TREEPSI [29]. 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.

4) Grid Based Data Aggregation Protocols

In grid-based data aggregation, a set of sensors is assigned as data aggregators in fixed regions of the sensor network called grids. The sensors in a particular grid transmit the data directly to the data aggregator of that grid. Data aggregator aggregates the data and transmits the aggregated data to sink node.

GROUP:

GROUP [19] was proposed by Liyang Yu, Neng Wang, Wei Zhang, Chunlei Zheng. In GROUP, 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.

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.

5) Hybrid Data Aggregation Protocols

Aggregation Tree Construction Based On Grid (ATCBG)

Jian Shu, Yebin Chen, Linlan Liu, Sheng Zhang and Jun Li proposed ATCBG [31]. 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. 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.

Tree-Clustered Data Gathering Protocol (TCDGP)

Gurpreet Singh Chhabra, Dipesh Sharma proposed TCDGP [30]. It is a combination of cluster-based and tree-based protocols. 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.

Clustered Diffusion with Dynamic Data Aggregation (CLUDDA):

A hybrid approach CLUDDA has been proposed by S. Chatterjea and P.Havinga.CLUDDA [12] combines directed diffusion with clustering during the initial phase of interest or query propagation. The clustering approach ensures that only cluster heads and gateway nodes which perform inter cluster communication are involved in the transmission of interest messages. This technique conserves energy, since the regular nodes remain silent unless they are capable of servicing a request. In CLUDDA, the aggregation points are dynamic. The data-aggregation task is not assigned to any specific group of nodes in the network. The nodes performing data aggregation change as the locations of source nodes change. Any cluster head or gateway node which has the knowledge of query definition can perform data aggregation. An interesting feature of CLUDDA is that a query cache is maintained at the cluster heads and gateway nodes. The query cache lists the different data components that were aggregated to obtain the final data. It also contains the addresses of neighboring nodes from which the data messages originated.

V. CONCLUSION

This paper provides the comparative study on various data aggregation techniques for wireless sensor networks. In this paper various data-aggregation algorithms in wireless sensor networks are elaborately surveyed and clearly explained. Also the comprehensive studies of various data aggregation protocols under the network architecture are discussed. The main features, advantages and disadvantages of various data aggregation algorithm are described clearly. 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.

ACKNOWLEDGMENT

The authors would like to acknowledge and thank Technical Education Quality Improvement Program [TEQIP] Phase 2, BMS College of Engineering and SPFU [State Project Facilitation Unit], Karnataka for supporting the research work.

REFERENCES

[1] Nandini. S. Patil, Prof. P. R. Patil, “Data Aggregation in Wireless Sensor Network”, IEEE International Conference on Computational Intelligence and Computing Research, 2010.

[2] Ramesh Rajagopalan And Pramod K. Varshney, “Data-Aggregation Techniques In Sensor Networks: A Survey”, IEEE Communications Surveys & Tutorials, 4th Quarter 2006

[3] Vimal Kumar, Sanjay Madria, Secure Hierarchical Data Aggregation in Wireless Sensor Networks: Performance Evaluation and Analysis”, IEEE 13th International Conference on Mobile Data Management, 2012

[4] J. Kulik, W.R. Heinzelman and H. Balakrishnan, “Negotiation-based protocols for disseminating information in wireless sensor networks,” Wireless Networks, vol. 8, March 2002, pp. 169-185.

[5] C. Intanagonwiwat, R. Govindan and D. Estrin, “Directed Diffusion: A Scalable and robust communication paradigm for sensor networks,” Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (MobiCOM '00), August 2000.

[6] Basavaraj S. Mathapati, Siddarama. R. Patil, “ Energy Efficient Reliable Data Aggregation Technique for Wireless Sensor Networks” International Conference on Computing Sciences, 2012.

[7] B. Krishnamachari, D. Estrin, and S. Wicker, “The impact of data aggregation in wireless sensor networks,” Proc. 22nd International Conference on Distributed Computing Systems Workshops, July 2002, pp. 575-78.

[8] B. Krishnamachari and J. Heidemann, “Application specific modeling of information routing in wireless sensor networks”, Proc. IEEE international performance, computing and communications conference, vol. 23, pp. 717-722, 2004.

[9] M. Thangaraj and P. Punitha Ponmalar, “A Survey on data aggregation techniques in wireless sensor networks”, International Journal of Research and Reviews in Wireless Sensor Networks (IJRRWSN) Vol. 1, No. 3, September 2011, ISSN: 2047-0037

[10] W. R. Heinzelman, “Application-specific protocol architectures for wireless networks”, PhD Thesis, Massachusetts Institute of Technology, June 2000.

[11] O. Younis and S. Fahmy, “HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks,” IEEE Transactions on Mobile Computing, vol .3 , no. 4, Dec 2004, pp. 366-79.

[12] S. Chatterjea and P.Havinga, “A Dynamic data aggregation scheme for wireless sensor networks,” Proc. Program for Research on Integrated Systems and Circuits, Veldhoven, The Netherlands, Nov. 2003.

[13] 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 March 2011 pp. 113-124

[14] 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.

[15] S. Lindsey, C. Raghavendra, and K.M. Sivalingam, “Data gathering algorithms in sensor networks using energy metrics,” IEEE Trans. Parallel and Distributed Systems, vol. 13, no. 9, September 2002, pp. 924-935.

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

[17] M. Ding, X. Cheng and G. Xue, “Aggregation tree construction in sensor networks,” *2003 IEEE 58th Vehicular Technology Conference*, vol.4, no.4, October 2003, pp 2168-2172.

[18] H. O. Tan and I. Korpeoglu, “ Power efficient data gathering and aggregation in wireless sensor networks,” *SIGMOD Record*, vol. 32, no. 4, December 2003, pp 66-71.

[19] 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

[20] Jun Yuea, Weiming Zhang, Weidong Xiao, Daquan Tang, Jiuyang Tang, “Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks” International Workshop on Information and Electronics Engineering (IWIEE), 2012.

[21] Jiao Zhang , Fengyuan Ren, Tao He , Chuang Lin, “Attribute-aware Data Aggregation Using Dynamic Routing in Wireless Sensor Networks”, IEEE 2010

[22] Jamal N. Al-Karaki, Raza Ul-Mustafa, Ahmed E. Kamal, “Data aggregation and routing in Wireless Sensor Networks: Optimal and heuristic algorithms” Science Direct, 2008

[23] Wenzhong Guo, Naixue Xiong, Athanasios V, Vasilakos, Guolong Chen, Hongju Cheng, “Multi-Source Temporal Data Aggregation in Wireless Sensor Networks”, springer, April 2010.

[24] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, “In-Network Aggregation Techniques for Wireless Sensor Networks: A Survey”, IEEE Wireless communication 2007.

[25] Braginsky D, Estrin D, “Rumor routing algorthim for sensor networks,” Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, Atlanta, GA, pp. 22-31, 2002.

[26] Manjeshwar, E.; Agrawal, D.P. TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks. In Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS), San Francisco, CA, USA, 23–27 April 2001; pp. 2009–2015.

[27] Ye, M.; Li, C.; Chen, G.; Wu, J. EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks. In Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference (IPCCC), Phoenix, AZ, USA, 7–9 April 2005; pp. 535–540.

[28] 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.

[29] 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 .

[30] 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

[31] 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

[32] Prakashgoud Patil, Umakant Kulkarni, “Analysis of Data Aggregation Techniques in Wireless Sensor Networks” IJCEM International Journal of Computational Engineering & Management, Vol. 16 Issue 1, January 2013