

Data Fusion and Data Aggregation/Summarization Techniques in WSNs: A Review

Sakshi Chhabra
Computer Science & Engineering
MIET,
Meerut, India

Dinesh Singh
Computer Science & Engineering
DCRUST, Murthal
Sonapat, India

ABSTRACT

Wireless sensor networks are energy constrained networks. Energy consumption in these networks can be reduced by processing the raw data at individual nodes through the application of suitable aggregation technique so that there is minimum amount of data that need to be transmitted towards the sink. The data aggregation functions that are applied should adhere to correctness, and should be computationally less complex considering the capabilities of the sensor nodes. In this paper, a brief survey on the present aggregation protocols and their impact, and some of the techniques that are applied at individual sensor nodes to reduce sensed data are presented.

Keywords

Wireless sensor networks, data fusion, data aggregation

1. INTRODUCTION

Wireless sensor networks (WSNs) [1],[2],[3] are networks with large number of wireless sensor nodes (SNs) deployed in an unattended mode. Some of the characteristics of WSNs include limited battery power and memory of SNs, single and multiple base stations, node dynamicity, data redundancy, no global unique ID, consistent topology change, single hop or multihop transmission, fault tolerant, etc. These networks have applications [4] ranging from military to industry, home automation and many more. The SNs in these networks are densely deployed and each one is equipped with basic units of sensing, computing and communicating. In these networks, SNs are placed in the close vicinity of the sources of events.

The data collected by these networks is significantly different from systems which rely on small number of highly sensitive elements placed sparsely [5]. The difference is due to the distributed processing applied on the sensed data. Hence, the sensor networks provide information with improved accuracy as the SNs in these networks can collaborate with each other to improve the quality of data. Other important feature of these networks includes self organization. Either SNs initialize by discovering their neighbors and building local area neighbor tables, or the base station learns network topology and creates spanning tree routing structure [6].

WSNs provide remote access to sensed data by providing sink nodes. These sink nodes act as a link to the outside world and connect the network to the other networks, such as internet. WSNs have various constraints like limited power, limited communication bandwidth, and limited memory. Since these networks are deployed in unattended environments, it becomes nearly impossible to replace the battery of the SNs once these are drained out. Thus the most critical constraint is the battery capacity and hence, the ultimate goal of WSNs

confines to reliable communication with least energy expenditure [6].

The energy of the network can be reserved by employing energy efficient mechanisms for communication and computing. Further, it has been observed that the energy consumed for communication is far greater than the energy required for computing. Therefore, it becomes extremely important to propose methodologies which include larger computations at SNs and lesser communication between them, thereby consuming lesser power. Various methods have been proposed to reduce the communication cost (in terms of energy). Firstly, the amount of data that needs to be transmitted across SNs can be reduced by aggregation; secondly, an energy efficient path for routing the messages can be traced and the third can be a mixture of these two techniques.

The remainder of this paper is organized as follows: Section 2 discusses the need of data mining in wireless sensor networks. In Section 3, terms relating to data mining process like data fusion and data aggregation are discussed. Section 4 highlights the data aggregation protocols, followed by Section 5 which discusses some of the processing techniques that are applied at individual nodes to reduce the data that needs to be forwarded. In Section 6, the impact of data aggregation is studied. Section 7 concludes the paper.

2. NEED OF DATA MINING PROCESS

In WSNs, the data sensed by the SNs can be destined to base station directly i.e. in a single hop or can be transferred via multiple hops. It is inefficient to communicate all the sensed data directly to the base station as there is minute difference in the sensed data of neighboring nodes. Further, this would lead to data redundancy at BS. Hence, sensed data must be aggregated or summarized at intermediate SNs by using appropriate data fusion mechanisms. Data gathering is defined as systematic collection of sensed data from multiple SNs to be eventually transmitted to base station for processing. These data gathering techniques must be energy efficient and balance load among different sensor nodes. Data fusion is not only helpful in avoiding data redundancy at BS, but also supports the network in the following manners:

- To help reduce collisions which would occur in case of huge network traffic, if every SNs is allowed to transmit raw data
- To avert the problem of false or malicious data from been sent by malicious SNs, as the data is sent after aggregation only.
- It assists in making useful inferences.

3. DATA FUSION AND RELATED TERMS

Quite often the terms data fusion and data aggregation are used interchangeably but there lies a difference between the two. Data fusion is defined as the use of techniques that combine data from multiple sources and gather this information in order to achieve inferences, correlations, associations which are more efficient and potentially more accurate than if they were achieved by means of a single source [7], whereas, the second term Data aggregation which is a subset of data fusion is just a process of summarizing the data coming from multiple SNs in order to reduce or eliminate redundant data.

Process of data fusion can be centralized or distributed [7]. In centralized data fusion techniques, all the sensed data is sent to a single SN which performs the task of fusion. On the contrary, in distributed techniques, the task of data fusion is not solely the responsibility of a single SN, but the work is distributed among several SNs. In distributed techniques, each SN performs fusion of its own data and data from neighboring nodes. Due to the battery scarcity of a SN, distributed data fusion techniques are preferred over centralized ones. Broadly, there are two prime motives to perform data fusion: one to improve accuracy of data and another to conserve energy.

Some of the applications of data fusion are to detect routing failures, to overcome sensor failures, improve location estimates of nodes, and overcome spatial and temporal coverage problems and to reduce energy consumption in the network. Besides data fusion and data aggregation, there are several other terms of importance such as sensor fusion and multisensor integration. Figure 1 elucidates the relationship between all these terms.

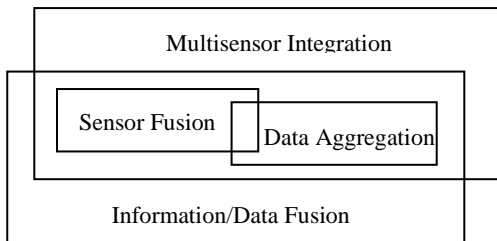


Figure 1: Relationship between different fusion terms

Sensor/Multisensor fusion is used to specify that the data coming from SNs is fused into one representational format. However, Multisensor integration is superset of sensor fusion, where in addition to fusing the data, it makes clear to the system that how this data can be used to accomplish a task by the system and interact with the environment. Data fusion can be categorized based on different parameters. Classification of the same is depicted in Fig. 2.

Data Aggregation

The basic idea of data aggregation is to aggregate data at certain SNs known as aggregators thereby eliminating redundancies and thus, reducing the number of transmissions between SNs. According to [7], “Data aggregation comprises the collection of raw data from pervasive data sources, the flexible, programmable composition of the raw data into less voluminous refined data, and the timely delivery of the refined data to data consumers”. Data aggregation reduces load on SNs, which helps in handling the data with priorities more effectively. Data aggregation techniques also have few disadvantages like increased delay, stronger hardware requirements, etc. [8].

Data aggregation shifts focus from address centric routing to data centric routing [9]. In address centric routing, individual SNs send data to the sink via a shortest path, whereas in data centric routing, different SNs send data to sink but routing nodes in between can perform aggregation on the data packets. The task of performing data centric routing with optimal data aggregation is a NP-hard. To reduce the complexity, three suboptimal schemes; Center at Nearest Source, Greedy Incremental Tree and Shortest Path Tree have been proposed [10].

The performance of aggregation methods depends on several factors like the number of source, their position in the network, the topology of the network, etc. Followings are several performance characteristics which are used to design data aggregation algorithms [11],[12]:

- Energy efficiency: Energy is the scarcest resource in WSN and must be consumed efficiently. A data aggregation scheme should be energy efficient. The ultimate goal of any data aggregation scheme is that each sensor possesses and spends the same amount of energy in each data gathering round.
- Network lifetime: Data aggregation schemes should balance the power consumption of all the nodes so that the network lifetime can be increased.
- Scalability: Aggregation protocols must be scalable enough so that they can easily and efficiently work with large number of nodes.
- Latency: Data aggregation mechanism should introduce least latency.
- Data accuracy: Data aggregation protocol should preserve the data accuracy and should also avoid compromised data.
- Overhead: An aggregation algorithm must be designed so as to minimize bandwidth utilization, power consumption and processing requirements.

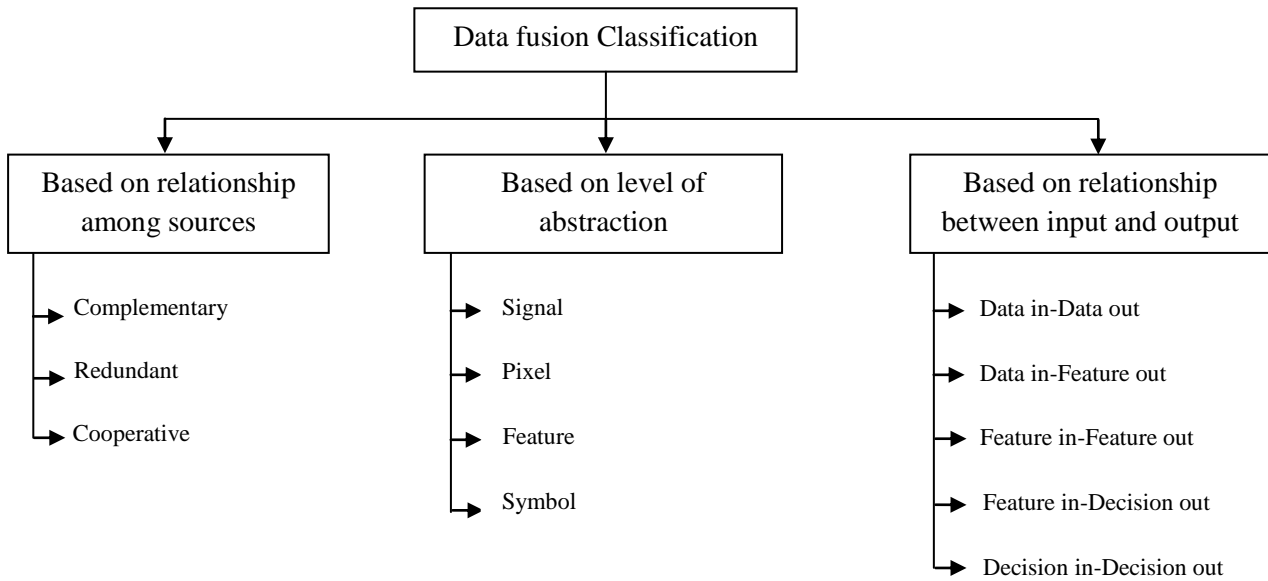


Figure 2: Classification of Data Fusion

4. DATA AGGREGATION PROTOCOLS AND THEIR CLASSIFICATION

Data aggregation protocols are proposed aiming at reducing the data needed to be transmitted to the base station. These protocols are broadly classified in two categories: Centralized/Server based and In-network approach. In centralized/server based schemes, all SNs readings are sent directly to the base station, which then computes the aggregates. There is excessive energy consumption of the nodes in this approach. This mode is particularly useful when base station is closer to the SNs and/or cost of receiving data is very high compared to the cost of transmitting.

The second approach is the In-network. Madden et al [13] proposed an in network Tiny AGgregation scheme (TAG). The scheme distributes the aggregation operators throughout the network and executes them in a power efficient manner. A tree is built where the user acts as a root and SNs act as leaves. The user sends queries, which are distributed throughout the network. SNs respond the data back towards the user. As the data flows back, it is aggregated by the intermediate nodes.

Classification of aggregation protocols

Aggregation protocols can be categorized based on tree structure, where the SNs are organized to form a tree. The sources generally act as leaf nodes and sink acts as a root. Data is aggregated at intermediate nodes from sources to sink. The hierarchy of these protocols is depicted in Fig. 3. Broader classification based on the tree structure is: planar and hierarchical protocols. The planar algorithms can be further categorized into: query routing based, chain routing based and suboptimal aggregation tree algorithms; while the hierarchical

algorithms are sub categorized as: cluster structure, cluster tree structure and cluster grid structure algorithms.

4.1 Planar Algorithms

In these algorithms, each sensor node plays the same role and has same capabilities. These algorithms include: query routing based algorithms, chain routing based algorithms, sub optimal aggregation tree algorithms.

4.1.1 Query Routing Based Algorithms

These algorithms follow data centric paradigm. Intanagonwiwat et al. [14] proposed directed diffusion, which is a query routing based protocol explained in Fig 4.

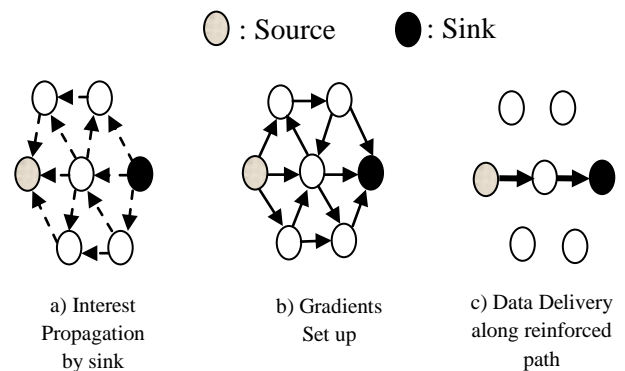


Figure 4: Directed Diffusion

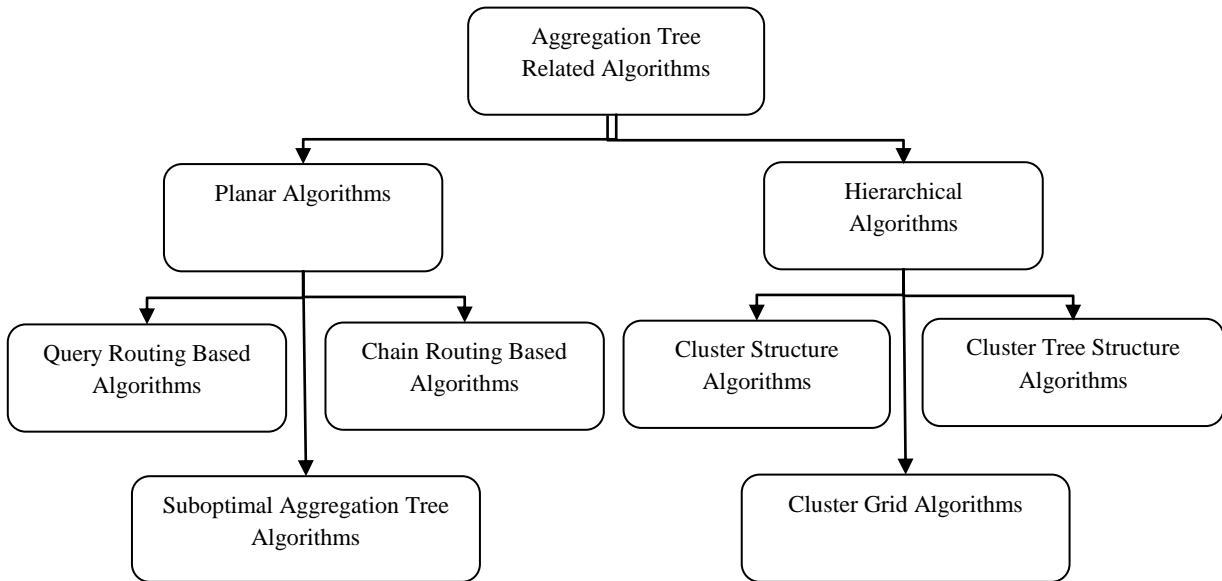


Figure 3: Hierarchy of Aggregation Tree Related Algorithms

Data in this protocol is named using attribute-value pairs. It consists of three phases: interest propagation & gradients set up, data delivery, and path reinforcement. An interest is propagated throughout the network by the sink. If the attributes of the data generated by source match the interest, a gradient is set up. The gradient decides the data rate and direction in which the data is sent. Data flows from sources to the sink along multiple paths. During path reinforcement, one of the paths is reinforced using certain criteria. One of the important features of directed diffusion is that interest propagation, data propagation and aggregation are determined locally through message exchange between neighboring SNs.

Kulik et al. [15] proposed Sensor Protocol for Information via Negotiation (SPIN). In this protocol, nodes name their data using meta-data. This meta-data is high level description of the data possessed by the node. Before transmitting data, nodes negotiate with each other to ensure that only useful information is transferred. The nodes also poll their resources

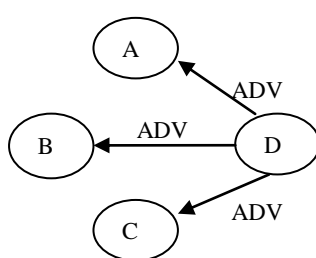
before any transmissions, which make them to restrain certain activities when their energy is low. The protocol consists of three types of messages for communication:

- ADV: These messages are used by the sensors to communicate if they have some new data to share and contain meta-data about the data.
- REQ: Nodes send this message when they want to receive some data.
- DATA: These messages contain the actual data sensed by the sensors.

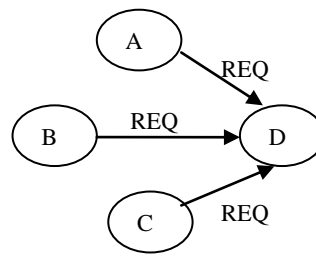
Figure 5 explains the working of the protocol. SPIN removes the three deficiencies of classic flooding: implosion, overlap and resource blindness via negotiation and resource adaption.

4.1.2 Chain Routing Based Algorithms

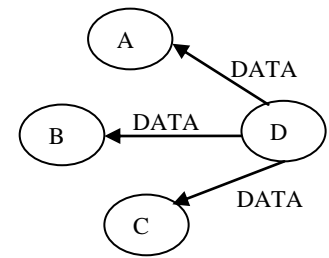
Lindsey et al. [16] proposed a chain based data aggregation protocol, PEGASIS.



a) Node D sending ADV messages to its neighbors



b) Neighbors sending REQ message back to D requesting for data



c) Node D sends DATA messages containing actual data

Figure 5: SPIN protocol

In this protocol, a chain of sensor nodes is constructed using greedy approach or the base station can create a chain and broadcast this to all the nodes. Nodes in the chain transmit data to their immediate neighbors. Each node except the end nodes is responsible for fusion of the data coming from its neighbors. Nodes take turns to become a leader node for e.g. node C in Fig. 6. This distributes the energy load among SNs, thereby balancing energy depletion among them.

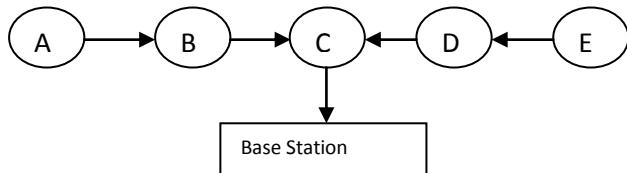


Figure 6: Node C sends aggregated data to base station

There also exists a chain based binary scheme for networks with CDMA nodes and a chain based 3-level scheme for networks with non-CDMA nodes in order to minimize energy*delay [17]. Figure 7 presents the binary scheme. A linear chain among all nodes is constructed. In every round, each node transmits to its close neighbor in a given level of hierarchy. The nodes which receive at a given level are the only active nodes in the next level. Finally, at the top level, the leader node is left which transmits data to the base station.

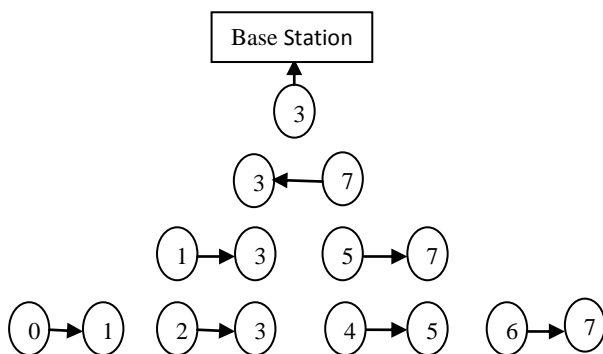


Figure 7: Chain Based Binary Scheme

As CDMA nodes are expensive, so a 3-level scheme for non-CDMA nodes is proposed, where the linear chain is divided into G groups of N/G number of successive nodes, where G is calculated based on the size of the network and the number of nodes in the network. One node from each group is active in the next level. Thus, there are G nodes in the next level which are again divided into two groups maintaining a three level hierarchy.

4.1.3 Sub Optimal Aggregation Tree Algorithms

As discussed previously, the task of optimal data aggregation is NP hard problem and therefore, sub optimal data aggregation algorithms have been proposed: CNS (Center at Nearest Source), GIT (Greedy Incremental Tree) and SPT (Shortest Path Tree).

- Center at Nearest Source (CNS): In this data aggregation scheme, all sources send data to a source which is nearest to the sink. This node then sends the aggregated data to the sink.

- Greedy Incremental Tree (GIT): At the start, the aggregation tree consists of the shortest path between sink and the nearest source. Later at each step, a source which is closest to current aggregation tree is added to the tree.
- Shortest Path Tree (SPT): In this scheme, each source sends its data to the sink along the shortest path between the two of them and the overlapping paths are combined to form a tree.

These three schemes are suitable for event driven WSN. When the sink node is far away, then GIT performs the best in terms of energy. When the sink node is very near, then CNS can't work effectively while the performance of other two depends on aggregation degree. If the aggregation degree is high, then GIT saves more energy than SPT.

4.2 Hierarchical Algorithms

Planar algorithms expend a lot of energy which reduce the network lifetime. Hence, hierarchical protocols were designed where data is aggregated at special intermediate nodes. This reduces the number of message transmissions to the sink which saves the energy of the network.

4.2.1 Cluster Structure Algorithms

In these algorithms the SNs are divided into groups called clusters as shown in Fig. 8. Each cluster has a leader called cluster head (CH). The CH's responsibility is intracuster coordination and intercluster communication. The SNs collect data from their surroundings and send it to their respective CHs. The data is then processed by the CH through the application of suitable data aggregation techniques and is forwarded to the sink. Since, only the processed data is forwarded, the communication cost is reduced which saves energy of the network.

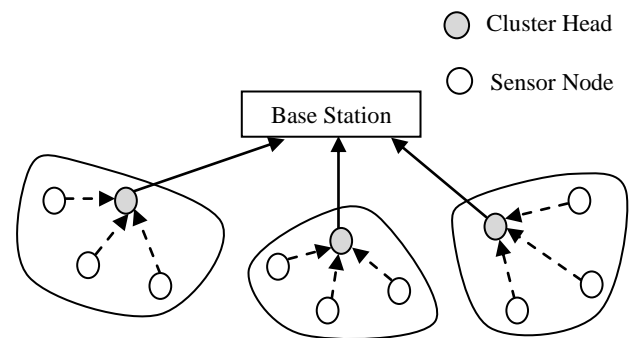


Figure 8: Cluster Based Approach

Nithyakalyani et al. [18] explains various questions that need to be considered in these type of aggregation approaches like how the CHs are elected, who initializes the CHs selection, when should the CHs be re-elected, when is a SN eligible to become a CH, etc.

Based on the distance between the CHs and their member nodes, there are two types of clustering strategies: single-level and multilevel clustering. In single-level clustering, the member nodes of the cluster send the data to their respective CHs in a single hop. The CHs then aggregate the data received from its member nodes and send it to the base station. The CHs acts as sink for member nodes and base station is the sink for the CHs. In multilevel clustering, a

clustering hierarchy is formed by replicating the above structure as is needed for the application.

The clustering based approaches are well known for their effectiveness, lower complexity and flexibility. These approaches reduce the energy consumption by improving bandwidth utilization, balance load among SNs, reduce data delay, reduce the number of SNs transmitting data to the base station and are also scalable.

Heinzelman et al. [19] proposed a distributed clustering based approach called LEACH (Low-energy Adaptive Clustering Hierarchy). The nodes do not require knowledge of the whole network in order for LEACH to operate. LEACH uses localized coordination and control for cluster set-up and operation. The operation is broken into rounds, where each round has an advertisement phase, cluster set-up phase, schedule creation and data transmission phase. Randomized rotation of CHs is done so that energy dissipation is distributed among all the SNs. Results show that LEACH reduces the communication energy by eight times as compared to direct transmission and minimum-transmission-energy routing.

Kumar et al. [20] proposed a protocol EECHDA (Energy Efficient Clustering Hierarchy and Data Aggregation) for homogenous WSN. Homogenous WSN are networks with SNs having same characteristics like same battery power, memory, etc. EECHDA extends the network lifetime by a factor of 50% as compared to direct transmission

4.2.2 Cluster Tree Structure Algorithms

In these algorithms, the network is first divided into set of clusters each with a CH, which is followed by the CHs forming a reverse multicast tree to transmit data to the sink, as shown in Fig. 9.

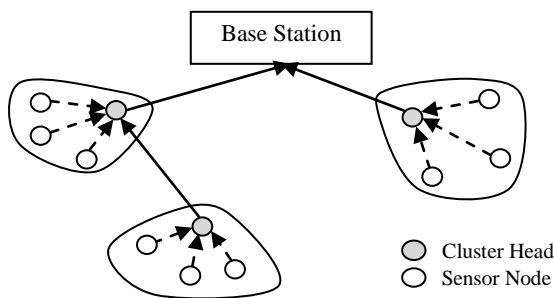


Figure 9: Cluster Tree Structure Algorithms

Here the data is sensed by the SNs and is transmitted to the CHs. The CHs then transmit the aggregated data to other CHs on the way to the sink

Xiao et al. [21] proposed a data fusion algorithm DFHCAC based on heterogeneous clustering. Different cluster have different number of SNs, so the energy consumption of each cluster is different and hence DFHCAC reelects CHs

asynchronously. The basic steps of DFHCAC are setting up of initial cluster, steady state phase, re-election of CHs and new round. DFHCAC uses multi-hop instead of single hop in order to save energy. Once the initial clusters are set up, cluster heads establish a minimum transmission path tree and a fusion path tree according to their distance to BS. Results demonstrate that survival time of DFHCAC is 47% longer than LEACH and energy consumption is 25% less than LEACH.

Younis et al. [22] presented a protocol HEED, a Hybrid, Energy Efficient Distributed clustering approach. The approach is hybrid because it uses a combination of node's residual energy and intra cluster communication cost to elect CHs. CHs can communicate with other CHs to aggregate their data and transfer data to sink via multiple hops. HEED incurs low message overhead and achieves uniform CHs distribution across the network.

4.2.3 Cluster Grid Algorithms

Yu et al. [23] proposed a Grid clustering routing protocol (GROUP). In this protocol, the network is assumed to have several sink nodes. A primary sink node is elected from these sink nodes which initiates the cluster grid formation. The network is divided into clusters and CHs are elected so that they form a grid. Each CH communicates with its member nodes and other CHs using different transmission power. The grid structure of CHs assures even cluster head distribution and makes routing between them simple. This protocol distributes the energy load among SNs and scales well.

Different research work exists related to classification of data aggregation protocols. Renjith et al [12] classifies aggregation into six categories: tree based data aggregation approach, multipath based data aggregation approach, cluster based data aggregation approach, hybrid data aggregation approach, centralized data aggregation approach and in-network data aggregation approach. Chen et al. [24] divides aggregation protocols into three categories: planar tree structure algorithms, cluster structure algorithms and cluster tree structure algorithms. Rajagoplan et al. [11] classifies data aggregation protocols for flat and hierarchical networks. Table 1 highlights some of the important points regarding some of these protocols.

It has been observed that the hierarchical protocols like the cluster based protocols decrease the energy consumption of the network. Hence these protocols must be considered to increase the network lifetime. Several other energy efficient data gathering protocols have been developed. Zhu et al. [25] proposed an Energy Efficient Routing Algorithm to Prolong Lifetime (ERAPL) which efficiently expends energy of the network. The algorithm builds a data gathering sequence (DGS), which avoids mutual transmissions and loop transmissions between nodes. The sequence gives for each SN, the SNs to which it can transmit data such that there is no formation of a loop. An outgoing traffic proportion (OTP) matrix is built using DGS and is optimized using genetic algorithms.

Table 1: Aggregation Protocols

Aggregation Protocols	Features				
	Model and Assumptions	Architecture	Organization Type	Advantage	Disadvantage
Directed Diffusion	Constant average density of sensor nodes, sinks are uniformly scattered across sensor field, no congestion occurs	Planar	Query Routing based	Robust, suitable for application with many sources and few sinks	Not suitable for application requiring continuous data delivery to sink
PEGASIS	Homogenous nodes, Base station is fixed at far distance from sensor nodes, All nodes are able to communicate directly with sink	Planar	Chain Based	No need of cluster formation, uses multihop routing, suitable for application where sink node is far from source node	Huge delay, necessity of global knowledge of all nodes' position to pick suitable neighbors and minimize maximum neighbor distance
LEACH	Base station is fixed and is located far from sensors, homogeneous and energy constrained nodes	Hierarchical	Cluster Structure	Simple, balances load, has little communication overhead, applicable in habitat monitoring and environment monitoring that require sensing the environment continuously	Uneven distribution of CHs, transmission of redundant data to base station
HEED	SNs are quasi-stationary and are homogenous in nature, SNs can control their transmission power level, links are symmetric	Hierarchical	Cluster Tree Structure	Fault tolerant, balances load, suitable for application like environmental monitoring which requires efficient data aggregation and prolonged network lifetime, scalable	Huge overhead in cluster formation, there exists the problem of hotspots
GROUP	Stationary SNs, SNs are aware of their location and can adjust transceiver power consumption	Hierarchical	Cluster Grid Structure	Scalable, balances load	Proper cell size must be determined in order to balance energy consumption and end to end delay

5. PROCESSING TECHNIQUES APPLIED AT INDIVIDUAL NODE

There exist various techniques which find an efficient data routing path from sources to sink where the cluster based approach reduces the energy consumption of the network to a great extent. The CH applies various data processing techniques like median, maximum, minimum, etc. on the data received from different sources in the cluster. This is done so that the CH has least amount of data to transmit further and reduce the communication cost (in terms of energy). In addition to these, there exist several other processing techniques that can be applied at the individual nodes in order to reduce the data sensed by the SNs. These may include techniques focusing on removal of redundant data, compression of data, etc. Some of the techniques that focus on data fusion are discussed in [26], [27].

Data aggregation at a particular node e.g. at CH can be done in two ways: with size reduction and without size reduction [28]. The former refers to combining and compressing the data received by a SN in order to reduce the size of the packet that need to be transmitted forward, while, the latter means combining the data packets received from different neighbors into a single packet without processing the value of the data.

Aggregates can be computed in various ways. One way to compute them is the Distributed Random Grouping (DRG) algorithm explained by Chen et al. [29]. The algorithm used is distributed, localized, simple, and robust and uses probabilistic grouping to converge to an aggregate value. There is no need of maintaining any global data structure. Each SN operates in three different modes: idle, leader and member mode. The algorithm works in rounds where in each round a SN becomes a leader with certain probability and invites neighbor SNs to join the group. A local aggregate value is computed by the group and is updated to all the members of the group. This process is repeated for several rounds and a correct global value of the aggregate is found. Results show that this algorithm works efficiently than gossip based algorithms.

The architecture that can be used for classification and recognition purposes in the field of data fusion is the neural network model. The reason for applying this is the similarity between data fusion and neural networks i.e. both carry out calculations and process large volume of data to get useful inferences. Sung et al. [30] explains these similarities and proposes a data fusion model based on neural network. The BP model proposed, uses clustering routing protocols such as LEACH. The model has three layers. The input and the first hidden layer reside in the member nodes of a cluster performing complicated applications in data processing, while the output layer and the second hidden layer reside in the cluster head performing simple applications. The SN acts as a neuron, the CH as an interneuron performing aggregate function and the overall network as a nervous system.

In [31], two architectures implementing ART1 and Fuzzy ART models at the CHs are proposed. In these architectures, wavelets are used for initial data processing of the sensory inputs and the neural network is used for categorization of the sensory inputs. The artificial neural network performs pattern clustering which leads to dimensionality reduction. The neural network algorithm is implemented on a platform of Smart-It units, which are kind of SNs. These Smart-It units only send

the cluster number where the current sensory input has been classified which reduces the data to be sent.

Other processing technique is explained by Singhal et al in [32], a simple median based sensor fusion function, D function. The sensors considered in the model are abstract sensors. These sensors read a physical parameter and give out an abstract interval estimate which is bounded and connected subset of real line. When the length (upper bound-lower bound) of all abstract sensors is same, the algorithm calculates the midpoint of each abstract sensor. Median of these midpoints is then calculated which gives the midpoint of interval to be estimated. Tolerance is added on both the sides of this midpoint so as to find the estimated interval. This function follows a similar technique to calculate median for varying length of abstract sensors. The D function satisfies the monotonicity property and lipschitz condition.

Bahi et al [33] proposed an aggregation scheme for periodic sensor networks using set similarity functions. It identifies duplicate sets of the data captured periodically. The method consists of two phases: one at the node level locally, and the other at the node at higher level i.e. the aggregator. In local aggregation, duplicate data measurements are not added to the set that is to be transmitted to the base station. At the aggregator level, similarities between neighboring SNs readings is identified and is aggregated using frequency filtering technique. Simulations showed that data integrity is preserved and only useful data is forwarded to the base station.

6. IMPACT OF DATA AGGREGATION

We have seen that data aggregation is useful to save the energy of WSN. With efficient data aggregation protocols like clustering protocols and data processing applied at individual nodes, huge number of transmissions can be reduced. Apart from increasing the lifetime of the network, data fusion also helps to make decisions that could not be made with the readings of an individual SN.

Krishnamachari et al [10] theoretically explains the impact of data aggregation. For address centric routing, the optimal number of transmissions (T_a) per datum is:

$$d_1 + d_2 + d_3 + \dots + d_n = \text{sum}(d_i)$$

where d_i is the shortest distance between source S_i and the sink, and n are number of sources.

The diameter (p) of a set of nodes is the maximum pairwise shortest paths between these nodes. If the source nodes have a diameter ≥ 1 , then the number of transmissions for data centric routing (T_d) has the following bounds:

$$T_d \leq (n-1)p + \min(d_i)$$
$$T_d \geq (n-1) + \min(d_i)$$

When $p < \min(d_i)$, then $T_a > T_d$. The paper shows that aggregation is very useful when there are many SNs and these SNs are far from sink.

Apart from this, there is also a delay associated with aggregation as in some cases data from near SNs is needed to be held back and wait for aggregation with data coming from farther SNs.

7. CONCLUSION

Wireless sensor networks are resource constrained networks. The main resource which makes them functional is their battery power, and therefore it becomes extremely important to save energy in these networks. This can be done by employing energy efficient aggregation protocols and by processing data at individual nodes. Though, these aggregation techniques save energy but there is always a trade off as these cause delay in the network. Depending on the tolerance level of delay in different applications, different aggregation techniques can be applied. An aggregation protocol that processes data at the node level and gathers data in an energy efficient manner can be built.

8. REFERENCES

- [1] Akyildiz, W. Su, Y. Sankarasubramaniam, and E.Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks: The International Journal of Computer and telecommunications Networking*, Vol. 38 Issue 4, 2002
- [2] F. Lewis, "Wireless Sensor Networks" *Smart environments: Technologies, Protocols and Applications*, pp.11-46, 2004
- [3] D. Puccinelli and M. Haenggi, "Wireless Sensor Networks: Applications and Challenges of Ubiquitous Sensing," *Circuits and Systems Magazine, IEEE*, Vol. 5, Issue 3, pp. 19-31, 2005
- [4] Y. Obaisat, R. Braun, "On Wireless Sensor Networks: Architectures, Protocols, Applications, and Management"
- [5] G. Pottie, "Wireless Sensor Networks," *Information Theory Workshop*, pp.139-140, 1998
- [6] J. Stankovic, "Wireless Sensor Networks," *Computer IEEE*, Vol. 41, Issue 10, pp. 92-95, 2008
- [7] A. Abdelgawad and M. Bayoumi, "Data Fusion In WSN," *Resource-Aware Data Fusion Algorithms for Wireless Sensor Networks, Lecture Notes in Electrical Engineering 118*, pp. 17-35, 2012
- [8] M. Vodel and W. Hardt, "Data Aggregation and Data Fusion Techniques In WSN/SANET Topologies - A Critical Discussion," *TENCON 2012- 2012 IEEE Region 10 Conference*, pp. 1-6, 2012
- [9] B. Krishnamachari, D. Estrin, and S.Wicker, "Modeling Data-Centric Routing in Wireless Sensor Networks.", *USC Computer Engineering Technical Report CENG02-14*, 2002.
- [10] B.Krishnamachari, D. Estrin and S.Wicker, "The Impact of Data Aggregation In Wireless Sensor Networks," *Proceeding ICDCSW '02 Proceedings of the 22nd International Conference on Distributed Computing Systems, IEEE*, pp. 575-578, 2002
- [11] R. Rajagopalan and P.Varshney, "Data Aggregation Techniques in Sensor Networks: A Survey," *IEEE Communication Surveys & Tutorials*, Vol. 8, pp. 48-63, 4th Quarter 2006
- [12] P.Renjith and E.Baburaj, "An Analysis of Data Aggregation In Wireless Sensor Networks," *International Conference on Radar, Communication and Computing (ICRCC), SKP Engineering College, Tiruvannamalai, TN., India*, pp. 62-71, 2012.
- [13] S. Madden, M. Franklin, J. Hellerstein, W. Hong, "TAG: a tiny aggregation service for Ad hoc sensor networks". *Boston : ACM Press*, pp. 131-146, 2002.
- [14] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," *Proc. 6th Annual Int'l. Conf. Mobile Comp. and Net. (MobiCOM '00)*, Aug. 2000.
- [15] J. Kulik, W. R. Heinzelman, and H. Balakrishnan, "Negotiation-based Protocols for Disseminating Information in Wireless Sensor Networks," *Wireless Networks*, vol. 8, Mar. 2002, pp.169–85.
- [16] S. Lindsey, C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information system". In *Proceeding of the IEEE Aerospace Conference. Montana: IEEE Aerospace and Electronic Systems Society*, pp. 1125-1130, 2002.
- [17] S. Lindsey, C. Raghavendra, K. Sivalingam, "Data gathering in sensor networks using the energy * delay metric". *Proceeding of the IPDPS Workshop on Issues in Wireless Networks and Mobile Computing'01*, pp. 23-27, 2001.
- [18] S. Nithyakalyani and S. Kumar, "Data Aggregation in Wireless Sensor Network Using Node Clustering Algorithms - A Comparative Study" *Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013)*
- [19] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks". In *Proceeding of 33rd Annual Hawaii Int'l Conf. on System Sciences. Maui : IEEE Computer Society*, pp. 3005-3014, 2000.
- [20] D. Kumar, T. Aseri and R. Patel, "EECHDA: Energy Efficient Clustering Hierarchy and Data Accumulation For Sensor Networks" *BIJIT - BVICAM's International Journal of Information Technology, Bharati Vidyapeeth's Institute of Computer Applications and Management, Vol. 2*, 2010
- [21] L. Xiao and Q. Liu, "A Data Fusion Using Un-even Clustering for WSN" *Advanced Intelligence and Awareness Internet (AIAI 2011)*, pp. 216-219, 2011
- [22] O. Younis and S. Fahmy, "HEED: a Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor networks," *IEEE Trans. Mobile Computing*, vol. 3, pp. 366–79, 2004
- [23] L. Yu, N. Wang, W. Xiang and C. Xheng, "GROUP: a Grid-clustering Routing Protocol for Wireless Sensor Networks," *Wireless Communications, Networking and Mobile Computing*, pp. 1-5, 2006

- [24] Y. Chen, J. Shu, S. Zhang, L. Liu and L. Sun, "Data Fusion In Wireless Sensor Networks," *Electronic Commerce and Security*, Vol. 2, pp. 504-509, 2009
- [25] Y. Zhu, W. Wu, J. Pan, Y. Tang, "An energy-efficient data gathering algorithm to prolong lifetime of wireless sensor networks," *Computer Communications*, Vol. 33, pp. 639–647, 2010
- [26] B. Beheshti and H. Michel, "Middleware/API and Data Fusion in Wireless Sensor Networks," *Systems, Applications and Technology Conference (LISAT)*, IEEE, pp.1-4, 2011
- [27] L. Li and L. Wei-jia, "The analysis of data fusion energy consumption in WSN," *2011 International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM)*, Vol. 1, pp. 310-313, 2011
- [28] N. Patil and P. Patil, "Data Aggregation in Wireless Sensor Network," *IEEE International Conference on Computational Intelligence and Computing Research*, 2010
- [29] J. Chen, G. Pandurangan and D. Xu, "Robust Computation of Aggregates in Wireless Sensor Networks: Distributed Randomized Algorithms and Analysis," *Information Processing in Sensor Networks*, IEEE, pp. 348 – 355, 2005.
- [30] W. Sung, Y. Liu, J. Chen and C. Chen, "Enhance the Efficient of WSN Data Fusion by Neural Networks Training Process" *Computer Communication Control and Automation (3CA)*, IEEE, vol. 2, pp. 373 - 376, 2010.
- [31] Kulakov, D. Davcev and G. Trajkovski "Application of Wavelet Neural-Networks in Wireless Sensor Networks," *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks*, IEEE, pp. 262 – 267, 2005 IEEE
- [32] D. Singhal and R. Garimella, "Simple Median Based Information Fusion in Wireless Sensor network" *International Conference on Computer Communication and Informatics*, 2012.
- [33] J. Bahi, A. Makhoul and M. Medlej, "Data Aggregation for Periodic Sensor Networks Using Sets Similarity Functions," *Wireless Communications and Mobile Computing Conference*, IEEE, pp. 559 – 564, 2011.