International Journal of Engineering Innovation & Research
Volume 1, Issue 2, ISSN: 2277 – 5668

Data Aggregation Techniques in Wireless Sensor Network: A Survey

Anindita Ray, Debashis De
Department of Computer Science & Engineering, West Bengal University of Technology
BF-142, Sector 1, Salt Lake City, Kolkata 700064, West Bengal, India.
dr.debashis.de@gmail.com

***Abstract*— Data aggregation is essential for the efficient operation of wireless sensor network. Data aggregation has been widely recognized as an efficient method to reduce energy consumption by reducing the number of packets sent. In this paper a comprehensive review of the existing data aggregation techniques in wireless sensor network have been presented. Suitable criterias have been defined to classify existing solutions. In this paper several open issues have been identified and discussed which propose directions for future research in this area.**

***Key Words*- LEACH, PEGASIS, SDAP, TAG.**

I. Introduction

The rapid advances in processor, memory, and radio technology have enabled the development of distributed networks of small, inexpensive sensor nodes that are capable of sensing, computation, and wireless communication [1], [2], [3], [4]. For ease of deployment, sensor devices should be inexpensive, small, and have a long lifetime, which makes it important to develop very efficient software and hardware solutions. For this reason, protocols for sensor networks should be carefully designed so as to make the most efficient use of the limited resources in terms of energy, computation, and storage. The area of communications and protocol design [6], [9], [41] for sensor networks has been widely researched in the past few years, and many solutions have been proposed and compared. *In this survey we focus on an important aspect of sensor network: data aggregation.* In-network data aggregation is at the heart of sensor network research which allows trading off communication for computational complexity. For a given application area, network resource constraints, local computation often consumes significantly less energy than communication. In particular, resource efficiency, timely delivery of data [16]-[18], [25] to the sink node, and accuracy of the results [47] are conflicting goals, and the optimal trade-off among them largely depends on the specific application. Data aggregation has been widely recognized as an efficient method to reduce energy consumption in wireless sensor networks, which can support a wide range of applications such as monitoring temperature, humidity, level, speed etc. Section II describes the basic concepts of data aggregation, section III presents the objectives of data aggregation, section IV describes the basic ingredients of In-network data aggregation, section V presents several routing protocols regarding data aggregation, section VI represents data representation and in-network data aggregation function, section VII presents challenges of data aggregation and few existing solutions and section VIII concludes this paper.

II. Basic Concepts of In-Network Data Aggregation

In typical sensor network scenarios, data is collected by sensor nodes throughout some area, and needs to be made available at some central sink node(s), where it is processed, analyzed, and used by the application. The data sampled by the same kind of sensors have much redundancy since the sensor nodes are usually quite dense in wireless sensor networks. To make data aggregation more efficient, the packets with the same attribute, defined as the identifier of different data sampled by different sensors such as temperature sensors, humidity sensors, etc., should be gathered together. Data aggregation techniques are tightly coupled with how data is gathered at the sensor nodes as well as how packets are routed through the network, and have a significant impact on energy consumption and overall network efficiency. In-network data aggregation can be considered a relatively complex functionality, since the aggregation algorithms should be distributed in the network and therefore require coordination among nodes to achieve better performance. Also, we emphasize that data size reduction through in-network processing shall not hide statistical information about the monitored event. Fig.1. shows the structure of data aggregation in multiple levels.

According to [8] the in-network aggregation process is defined as follows: In-network aggregation is the global process of gathering and routing information through a multihop network, processing data at intermediate nodes with the objective of reducing resource consumption, thereby increasing network lifetime. In [8] it defines two approaches:

- *In-network aggregation with size reduction:* In this approach it combines and compresses data coming from different sources in order to reduce the information to be sent over the network. As an example, assume that a node receives two packets from two different sources containing the locally measured temperatures. Instead of forwarding the two packets, the sensor may compute the average of the two readings and send it in a single packet.
- *In-network aggregation without size reduction:* In this approach it merges packets coming from different sources into the same packet without data processing: assume receiving two packets carrying different physical quantities (e.g., temperature and humidity). These two

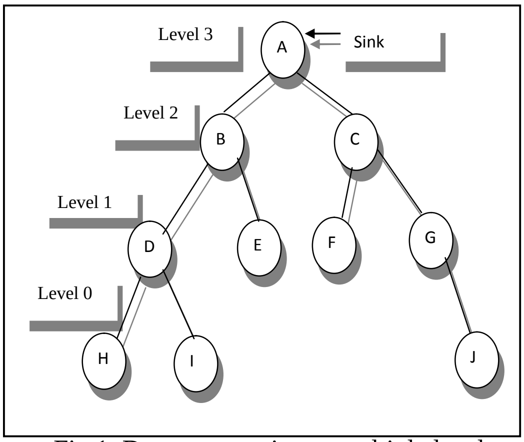

Fig.1. Data aggregation at multiple levels

values cannot be processed together, but they can still be transmitted in a single packet, thereby reducing overhead. The first approach is better able to reduce the amount of data to be sent over the network, but it may also reduce the accuracy with which the gathered information can be recovered at the sink. After the aggregation operation, it is usually not possible to perfectly reconstruct all of the original data. This actually depends on the type of aggregation function in use (i.e., lossy or lossless.)The second approach instead preserves the original information. Which solution to use depends on many factors including the type of application, data rate, network characteristics, and so on.

III. Objectives Of In-Network Data Aggregation

Main objectives of Data aggregation are to reduce energy consumption by reducing redundant packet and to increase network life time:

To reduce energy consumption by reducing redundant packet:

The duplicated packets in transmission are of great size in WSNs, because the readings from nodes can be similar due to the real environment [44]. But these replicated data cannot be simply discarded, since they may act important role in improving the accuracy of monitoring. Adaptive Data Aggregation Mechanism (ADAM) is applied here to utilize a sequence number to represent these repeated portions of data so as to reduce the total data transmitted and guarantee the QoS [45]. Data sampled by the same kind of sensors have much redundancy since the sensor nodes are usually quite dense in wireless sensor networks. To make data aggregation more efficient, the packets with the same attribute, defined as the identifier of different data sampled by different sensors such as temperature sensors, humidity sensors, etc., should be gathered together. In Attribute-aware Data Aggregation mechanism using Dynamic Routing (ADADR), it introduce packet attribute into data aggregation which can make packets with the same attribute convergent as much as possible and therefore improve the efficiency of data aggregation [46]. In [49] it uses measurement co-occurrence to identify data redundancy and a novel collaborative data gathering approach utilizing co-occurrence that offers a trade-off between the communication costs of data gathering versus errors at estimating the sensor measurements at the base station.

Increase Network Life Time:

A wireless sensor network consisting of a large number of nodes with limited battery power should minimize energy consumption at each node to prolong the network lifetime. In [20] it presents an exact as well as approximate algorithm to find the minimum number of aggregation points in order to maximize the network lifetime. In [27] a routing protocol based on a balanced tree, called BATR (Balanced Aggregation Tree Routing) is proposed, which uses near optimal minimal spanning tree for balancing the power consumptions over all nodes. The key idea is that a near optimal data aggregation can be achieved in terms of network lifetime when power consumption at each node can be nicely balanced. In Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs) [28] a new energy-efficient routing approach is presented that combines the desired properties of minimum spanning tree and shortest path tree-based routing schemes. The proposed scheme uses the advantages of the powerful localized structures such as RNG and LMST and provides simple solutions to the known problems in route setup and maintenance because of its distributed nature. The proposed algorithm is robust, scalable, and self-organizing. The algorithm is appropriate for systems where all the nodes are not in direct communication range of each other.

IV. Basic Ingredients of In-Network Aggregation

Three basic ingredients of In-network aggregation techniques are:

- suitable routing protocols,
- effective aggregation functions,
- efficient ways of representing the data.

In the remainder of this section we briefly introduce each of these aspects.

A. Routing Protocols:

Routing protocols [5], [6], [9], [15], [41] plays important roles in data aggregation. Data aggregation requires a different forwarding paradigm than classic routing. Classic routing protocols typically forward data along the shortest path to the destination. However in time of data aggregation energy expenditure should be minimized, nodes should route packets based on packet content and the next hop should be chosen in order to promote in-network aggregation. This type of data forwarding is often referred to as data-centric routing. According to the data-centric paradigm, as a node searches for relay nodes, it needs to use metrics that take into account the positions of the most suitable aggregation points, the data type, the priority of the information, and so on.

B. Aggregation functions:

One of the most important functionalities that in-network aggregation techniques should provide is the ability to combine data coming from different nodes. There are several types of aggregation functions [13]–[20], and most of them are closely related to the specific sensor application. Nevertheless, we can identify some common paradigms for their classification:

- *Lossy and lossless:* Aggregation functions can compress and merge data according to either a lossy or a lossless approach. In the first case the original values cannot be recovered after having merged them by means of the aggregation function. In addition, we may lose precision with respect to transmitting all readings uncompressed. In contrast, the second approach (lossless) allows us to compress the data by preserving the original information. This means that all readings can be perfectly reconstructed from their aggregate at the receiver side.
- *Duplicate sensitive and duplicate insensitive:* An intermediate node may receive multiple copies of the same information. In this case, it may happen that the same data is considered multiple times when the information is aggregated. If the aggregation function in use is duplicate sensitive, the final result depends on the number of times the same value has been considered. Otherwise, the aggregation function is said to be duplicate insensitive. For instance, a function that takes the average is duplicate sensitive, whereas a function that takes the minimum value is duplicate insensitive.

Good aggregation functions for wireless sensor networks should also consider additional requirements like very limited processing and energy capabilities of sensor devices, and should therefore be implementable by means of elementary operations. Also, different devices may be suitable for different types of operations, depending on their energy resources and computation capabilities. These facts need to be considered in the design of aggregation functions and routing protocols.

C. Data Representation:

Due to its limited storage capabilities, a node may not be able to store all the received/generated information in its internal buffer. It therefore needs to decide whether to store, discard, compress, or transmit the data. All these operations require a suitable way to represent the information [21]–[24]. The corresponding data structure may vary according to the application requirements. Finally, even though the data structure is usually common to all nodes, it should be adaptable to node-specific or location specific characteristics. A recent and promising method to deal with data representation and compression is distributed source coding techniques that compress data on the basis of some knowledge about its correlation. Although we described routing, aggregation, and data representation in isolation, they are intimately related and should be designed and implemented jointly for optimal performance.

V. Networking Protocols and Hierarchies for In-Network Aggregation

Many studies proposed solutions exploiting tree-based (or hierarchical) structures [5]-[7], [14], [27] in order to facilitate the in-network aggregation. These consist of routing algorithms based on a tree rooted at the sink. Sometimes the tree structure can be optimized to the type of data to be gathered. Also, the nodes can be locally grouped into clusters for improved efficiency. Some papers address the weaknesses of the tree-based approach by focusing on multipath routing. Finally, some very recent schemes implement a mixture of tree-based and multipath solutions referred as hybrid approaches to emphasize the adaptive nature of their routing algorithms.

In the following, we focus on each class of routing protocols separately (tree-based, cluster based, multipath, and hybrid) by reviewing the main concepts and briefly commenting on the pros and cons of each scheme.

A. Tree based approaches:

According to the survey of Routing protocols [42], [43] classic routing strategies are usually based on a hierarchical organization of the nodes in the network. In fact, the simplest way to aggregate data flowing from the sources to the sink is to elect some special nodes that work as aggregation points and define a preferred direction to be followed when forwarding data. In addition, a node may be marked as special depending on many factors such as its position within the data gathering tree [10], its resources [11], the type of data stored in its queue [12], or the processing cost due to aggregation procedures [13]. According to the tree-based approach [5]-[7], [27] a spanning tree rooted at the sink is constructed first. Subsequently, such a structure is exploited in answering queries generated by the sink. This is done by performing in-network aggregation along the aggregation tree by proceeding level by level from its leaves to its root. Thus, as two or more messages get to a given node, their aggregate can be computed exactly. There are some studies where the sink organizes routing paths to evenly and optimally distribute the energy consumption while favoring the aggregation of data at the intermediate nodes [12], [14]. In [19] the authors compute aggregation by taking into account the residual energy of each node through linear programming. In [20] the authors investigate which nodes in the network can be exploited as aggregation points for optimal performance. In [10],[21] the focus is on the nodes that should be entrusted with the transmission of the sensed values, whereas in [11] the emphasis is put on the proper scheduling of sleeping/active periods. Often, optimal paths are calculated in a centralized manner at the sink by exploiting different assumptions on the data correlation and selecting the best aggregation points by means of cost functions [22]. BATR (Balanced Aggregation Tree Routing), which uses near optimal minimal spanning tree for balancing the power consumptions over all nodes [27]. The key idea is that a near optimal data aggregation can be achieved in terms of network lifetime when power consumption at each node can be nicely balanced. Recently, tree-based schemes for real-time or time-constrained applications have also been proposed [23]–[25]. Finally, a last approach based on aggregation trees relies on the construction of connected dominating sets [26]. These sets consist of a small subset of nodes that form a connected backbone and whose positions are such that they can collect data from any point in the network. Nodes that do not belong to these sets are allowed to sleep when they do not have data to send. Some rotation of the nodes in the dominating set is recommended for energy balancing. In the following paragraphs some of the main routing approaches based on aggregation trees have been reviewed.

TAG — The *Tiny Aggregation* (TAG) approach [15] is a data-centric protocol. It is based on aggregation trees and specifically designed for monitoring applications. This means that all nodes should produce relevant information periodically. Therefore, it is possible to classify TAG as a periodic per hop adjusted aggregation approach. The implementation of the core TAG algorithm consists of two main phases:

• The distribution phase, where queries are disseminated to the sensors

• The collection phase, where the aggregated sensor readings are routed up the aggregation tree.

For the distribution phase, TAG uses a tree based routing scheme rooted at the sink node. The sink broadcasts a message asking nodes to organize into a routing tree and then sends its queries. Each message contains a field specifying the level, or distance from the root, of the sending node. Whenever a node receives a message and it does not yet belong to any level, it sets its own level to be the level of the message plus one. It also elects the node from which it receives the message as its parent. The parent is the node that is used to route messages toward the sink. Each sensor then rebroadcasts the received message adding its own identifier (ID) and level. This process continues until all nodes have been assigned an ID and a parent. The routing messages are periodically broadcast by the sink in order to keep the tree structure updated. After the construction of the tree, the queries are sent along the structure to all nodes in the network.TAG adopts the selection and aggregation facilities of the database query languages (SQL).Accordingly, TAG queries have the following form:

SELECT{agg(expr), attrs} from SENSOR
WHERE{selPreds}
GROUP BY{attrs}
HAVING{havingPreds}
EPOCH DURATION i

In practice, the sink sends a query, where it specifies the quantities that it wants to collect (attrs field), how these must be aggregated (agg (expr)), and the sensors that should be involved in the data retrieval. An EPOCH DURATION field specifies the time (in seconds) each device should wait before sending new sensor readings. This means the readings used to compute an aggregate record all belong to the same time interval, or epoch. During the data collection phase, due to the tree structure, each parent has to wait for data from all of its children before it can send its aggregate up the tree. TAG may be inefficient for dynamic topologies or link/device failures. In addition, as the topology changes, TAG has to reorganize the tree structure, which means high costs in terms of energy consumption and overhead.

Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs)

Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs) is based on topologies, such as LMST and RNG that can approximate minimum spanning tree and can be efficiently computed using only position or distance information of one-hop neighbors [28]. The actual routing tree is constructed over these topologies. It also considers different parent selection strategies while constructing a routing tree. The solution involves route maintenance procedures that will be executed when a sensor node fails or a new node is added to the network. The proposed solution is also adapted to consider the remaining power levels of nodes in order to increase the network lifetime.

Directed Diffusion: Directed Diffusion [5] is a reactive data-centric protocol. The routing scheme is specifically tailored for those applications where one or few sinks ask some specific information by flooding the network with their queries. Directed Diffusion is organized in three phases

• Interest dissemination
• Gradient setup
• Data forwarding along the reinforced paths

When a certain sink is interested in collecting data from the nodes in the network, it propagates an *interest* message (*interest dissemination*), describing the type of data in which the node is interested, and setting a suitable operational mode for its collection. Each node, on receiving the interest, rebroadcasts it to its neighbors. In addition, the node sets up *interest gradients*, that is, vectors containing the next hop that has to be used to propagate the result of the query back to the sink node (*gradient setup*) as shown in Fig.2. If the sink sends an interest that reaches nodes *a* and *b*, and both forward the interest to node *c*, node *c* sets up two vectors indicating

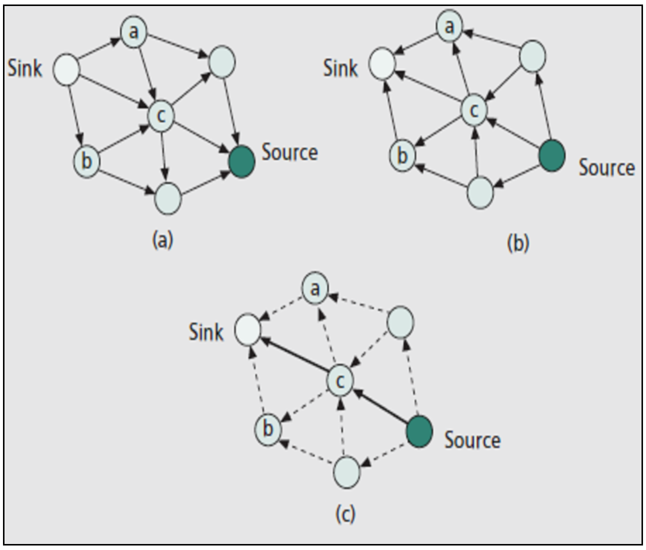

Fig.2. A simplified scheme for Directed Diffusion [5]: a) interest dissemination; b) gradients setup; c) data delivery along the reinforced path.

that the data matching that interest should be sent back to *a* and/or *b*.

The strength of such a gradient can be adapted, which may result in a different amount of information being redirected to each neighbor. To this end, various metrics such as the node's energy level, communication capability, and position within the network can be used. Each gradient is related to the attribute for which it has been set up. As the gradient setup phase for a certain interest is complete, only a single path for each source is *reinforced* and used to route packets toward the sink. Data aggregation is performed when data is forwarded to the sink by means of proper methods, which can be selected according to application requirements.

The data gathering tree must be periodically refreshed by the sink, and this can be expensive in dynamic topologies. A trade-off, depending on the network dynamics, is involved between the frequency of the gradient setup and the achieved performance. A valuable feature of Directed Diffusion consists of the *local interaction* among nodes in setting up gradients and reinforcing paths. This allows for increased efficiency as there is no need to spread the complete network topology to all nodes in the network. Several approaches [29], [30] have been proposed to reduce the control traffic generated by the local interactions among nodes with Directed Diffusion where the authors use properly defined aggregation trees with the main purpose of reducing both traffic and delay. In [29] a modified version of Directed Diffusion, *Enhanced Directed Diffusion* (EDD), is proposed.

PEGASIS: The key idea in Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [31] is to organize the sensor nodes in a chain. Moreover, nodes take turns acting as the chain leader, where at every instant the chain leader is the only node allowed to transmit data directly to the sink. In this way it is possible to evenly distribute the energy expenditure among the nodes in the network. The chain can be built either in a centralized (by the sink) or distributed manner (by using a greedy algorithm at each node).In both cases, however, the construction of the chain requires global knowledge of the network at all nodes. The chain building process starts with the node furthest from the sink. Then the closest neighbor to this node is chosen as the next one in the chain, and so on. Nodes take turns acting as leader according to the following rule: Node i is elected as the leader in round i. If there are N nodes in the network, rounds cyclically take values in {1, 2, …, N}according to a TDMA schedule. As a consequence, the leader is not always the same, but during each transmission round it is at a different position in the chain.

In PEGASIS, each node receives data from a neighbor and aggregates it with its own reading by generating a single packet of the same length. Subsequently, such an aggregate is transmitted to the next node in the chain until the packet reaches the current chain leader. At this point the leader includes its own data into the packet and sends it to the sink. Figure 3 shows the chain-based data-

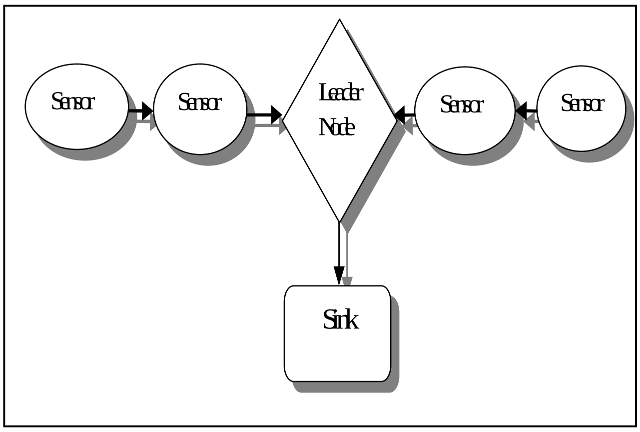

Fig.3. Chain based data aggregation in PEGASIS

aggregation procedure in PEGASIS where the arrows indicate the direction of data transmission.

The main disadvantages of PEGASIS are the necessity of having a complete view of the network topology at each node for proper chain construction and that all nodes must be able to transmit directly to the sink. This makes the scheme unsuitable for those networks with a time varying topology. In addition, link failures and packet losses may also affect the performance of this protocol. In fact, the failure of any intermediate node compromises the delivery of all data aggregated and sent by the previous nodes in the chain. Hence, some improvements to the scheme may be needed in order to increase its robustness.

Drawbacks of Tree Based Approaches

Tree Based Approaches has some drawbacks as actual wireless sensor networks are not free from failures. More precisely, when a packet is lost at a given level of the tree, the data coming from the related sub tree are lost as well. In fact, a single message at a given level of the tree may aggregate the data coming from the whole related sub tree. In spite of the potentially high cost of maintaining a hierarchical structure in dynamic networks and the scarce robustness of the system in case of link/device failures, these approaches are particularly suitable for designing optimal aggregation functions and performing efficient energy management.

B. Cluster Based Approaches

Similar to tree-based algorithms, cluster-based schemes [29], [32], [33], [34] also consist of hierarchical organization of the network. However, here nodes are subdivided into clusters. Moreover, special nodes, referred to as cluster heads, are elected in order to aggregate data locally and transmit the result of such aggregation to the sink. The advantages and disadvantages of cluster-based schemes are very similar to those of tree-based approaches.

LEACH: Low Energy Adaptive Clustering Hierarchy (LEACH) [9] is the first hierarchical cluster-based routing protocol for wireless sensor network which partitions the nodes into clusters, in each cluster a dedicated node with

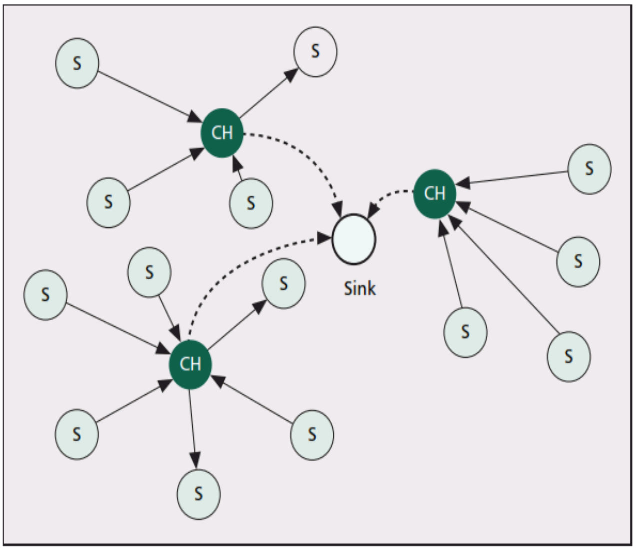

Fig. 4. LEACH Clustering Approach [9]

extra privileges called Cluster Head (CH) is responsible for creating and manipulating a TDMA schedule and sending aggregated data from nodes to the base station (BS) where these data is needed using CDMA. Remaining nodes are cluster members as shown in Fig.4.

In LEACH cluster-heads are stochastically selected. In order to select cluster-heads each node *n* determines a random number between 0 and 1. If the number is less than a threshold *T (n)*, the node becomes a cluster-head for the current round. The threshold is set as follows:

$$T(n) = \frac{p}{1 - p \times \left(r \bmod \frac{1}{p} \right)} \quad \forall\, n \in G$$

$$T(n) = 0 \quad \forall\, n \notin G \qquad (1)$$

where P is the cluster-head probability, r is the number of the current round and G is the set of nodes that have not been cluster-heads in the last 1/P rounds. This algorithm ensures that every node becomes a cluster-head exactly once within 1/P rounds.

In LEACH-C, an enhancement over the LEACH protocol was proposed [41]. This protocol uses a centralized clustering algorithm and the same steady-state phase as LEACH. LEACH-C protocol can produce better performance by dispersing the cluster heads throughout the network. In [55] it presents a clustering hierarchy protocol based on energy efficient cluster head selection (EECHS) which aims to reduce energy consumption within the wireless sensor network and prolong the lifetime of the network**.** In [56] it presents an improved version of LEACH protocol which aims to reduce energy consumption within the wireless sensor network and prolong the lifetime of the network. It improves LEACH protocol by improving the election strategy of the cluster-head nodes.

Cougar: Cougar [35] is most suitable for monitoring applications, where nodes produce relevant information periodically. The protocol can be classified as a periodic per hop aggregation approach. Cougar is basically a clustering scheme. As soon as the cluster heads receive all data from the nodes in their clusters, they send their partial aggregates to a gateway node. Of course, being similar to LEACH, Cougar is also affected by the same problems in highly dynamic environments. Unlike in LEACH, where each node picks its cluster head based on signal strength measurements, in Cougar cluster head selection may be driven by additional metrics. In fact, a node could be more than one hop away from its cluster head. For this reason, the routing algorithm adopted to exchange packets within clusters is based on the Ad Hoc On Demand Distance Vector (AODV) technique. As AODV does not generate duplicate data packets, Cougar is particularly suitable to perform in-network aggregation with duplicate sensitive aggregators. The core Cougar algorithm consists of the node synchronization engine, which ensures that data is aggregated correctly. Each cluster head has a waiting list containing all nodes from which it expects a message. The list is updated every time the node receives a record from a node in its cluster.

A major advantage of a clustered structure is that it directly allows aggregation of data at the cluster head. Such algorithms work well in relatively static networks where the cluster structure remains unchanged for a sufficiently long time, but they may be fragile when used in more dynamic environments.

C. Multipath Approaches

In order to overcome the robustness problems of aggregation trees, a new approach was recently proposed [36], [37]. Instead of having an aggregation tree where each node has to send the partial result of its aggregation to a single parent, these solutions send data over multiple paths. The main idea is that each node can send the data to its multiple neighbors by exploiting the broadcast characteristics of the wireless medium. Hence, data may flow from the sources to the sinks along multiple paths, and aggregation may be performed by each node. In contrast to the tree-based schemes discussed above, multipath approaches allow duplicates of the same information to be propagated. Clearly, such schemes trade higher robustness for some extra overhead due to sending duplicate packet. An aggregation structure that fits well with this methodology is called ring topology, where sensor nodes are divided into several levels according to the number of hops separating them from the data sink. Data aggregation is performed over multiple paths as packets move level by level toward the sink.

Synopsis Diffusion: In Synopsis Diffusion [36] Data aggregation is performed through a multipath approach. The underlying topology for data dissemination is organized in concentric rings around the sink. Synopsis Diffusion consists of two phases:

- The distribution of the queries
- The data retrieval phase

The ring topology is formed when a node sends a query over the network.

D. Hybrid Data Aggregation Approaches

In order to benefit from the advantages of both tree-based and multipath schemes, it is possible to define hybrid approaches that adaptively tune their data aggregation structure for optimal performance. To the best of our knowledge, a single work [37] has been proposed with this aim. The related protocol is presented next.

Tributaries and Deltas: The Tributaries and Deltas protocol [37] tries to overcome the problems of both tree and multipath structures by combining the best features of both schemes. The result is a hybrid algorithm where both data aggregation structures may simultaneously run in different regions of the network. The idea is that under low packet loss rates, a data aggregation tree is the most suitable structure due to the possibility of implementing efficient sleeping modes and the good efficiency in representing and compressing the data. On the other hand, in case of high loss rates or when transmitting partial results that are accumulated from many sensor readings, a multipath approach may be the best option due to its increased robustness. Hence, nodes are divided into two categories: nodes using a tree-based approach to forward packets (also called T nodes) and nodes using a multipath scheme (M nodes). This means that the network is organized in regions implementing one of the two schemes. The main difficulty is to link regions running different data aggregation structures. For that, edge correctness and Path correctness rules [37] have to be satisfied. Comparisons of different characteristics of routing protocols are shown in Table I.

VI. Data Representations And In-Network Aggregation Functions

The problems of finding proper data representation and an optimal aggregation function are strongly related and complex. The solutions proposed so far mostly adopt very simple aggregation functions such as average, median, min, and max [15], [31]. These strongly reduce the amount of data to be transmitted over the network but also heavily affect the precision of the transmitted information (lossy aggregation functions). However, in many cases we may be interested in a more detailed representation of the data, which calls for more complex functions and data structures which takes into account the spatial [51], temporal [52] correlation of the readings: cross-layer and self-adaptable data fusion rules have been proposed in [38], [39], [40]. In [51] simple data aggregation functions take into account the spatial correlation. In this strategy the dependence on the distance among nodes is quantified by a decay function which may, for example, decay exponentially with an increasing hop distance. During the data aggregation, each reading is weighed by a decaying factor that decreases with the distance to its source.

TiNA : Temporal coherency-aware in-Network Aggregation(TiNA) [52] take into account the temporal correlation in a sequence of sensor readings to reduce energy consumption by suppressing those values that do not affect the expected quality of the aggregated data. This is implemented through a TOLERANCE clause added to the SQL query. The tct parameter of this clause is used to specify the temporal coherency tolerance for the query.

Table I: Comparison of characteristics of different routing protocols

Routing Protocols	Characteristics			
	Organization Type	Energy saving Technique	Scalability	Overhead to maintain aggregation structure
LEACH	Cluster Based, Distributed	Rotation of the cluster heads	Low	Medium
COUGAR	Cluster Based, Distributed, Synchronous	Local route repairs	Low	Medium
PEGASIS	Chain-based	Rotation of the leader	Very Low	High
TAG	Tree-based, , driven by the sink	Sleeping Periods	Low	High
Directed Diffusion	Tree-based, , driven by the sink	None	Medium	High
Synopsis Diffusion	Multipath based, Distributed	None	High	Medium
Tributaries and Deltas	Tree/multipath based, driven by the sink	None	Medium	Medium

As an example, at a leaf node, each new available value ,V, is compared against the last reported data point; V_{new} is transmitted and aggregated up the tree if and only if it satisfies the following requirement (data suppression rule):

$$\frac{|V_{new} - V_{old}|}{V_{new}} > tct \quad (2)$$

TiNA uses the clause GROUP BY of the SQL query to decide how different messages shall be processed. The data gathering procedure executed at the internal nodes is as follows. They first gather and combine packets sent by their children. If a given node does not receive valid data from any of its children, it replaces the missing information using the last reported data from the same child (previously stored in its buffer). The node then considers its own reading. If it can be aggregated with some other data in its buffer, the reading is aggregated with that data regardless of the tct value. Doing so, internal nodes can report their values more often than leaf nodes, thus increasing the accuracy of the aggregation. On the other hand, if the internal node needs to create a new group, it does so and adds the new reading only if this data satisfies (2). The idea is that new groups are created only when the new measurements significantly differ from old data points.

AIDA: In AIDA [38] an aggregation scheme has been proposed that adaptively performs application-independent data aggregation (AIDA) in a time sensitive manner. It isolates aggregation decisions into a module that resides between the network and data link layers. AIDA performs lossless aggregation by concatenating network units into larger payloads that are sent to the MAC layer for transmission The AIDA architecture consists of a functional unit that aggregates and de-aggregates network packets as shown in Fig. 5. In addition, there is a control unit that adaptively controls timer settings and tunes the degree of aggregation. The transmission and control overhead is reduced by aggregation of multiple network units into a single AIDA aggregate.

Several versions of AIDA have been proposed ranging from aggregation decisions based on static thresholds to a dynamic online feedback control mechanism. In the fixed aggregation scheme, AIDA aggregates a fixed number of network units into an AIDA packet. In the on-demand aggregation scheme, AIDA-layer data aggregation takes place only when the MAC layer is available for transmission.

DADMA: In Data Aggregation and Dilution by Modulus Addressing (DADMA) [58] it proposes a distributed data aggregation and dilution technique for sensor networks where nodes aggregate or dilute sensed values according to the rules given in an SQL statement. In DADMA a wireless sensor network is treated as a distributed relational database. Database has a single view that is created by joining records which are locally stored in the sensor nodes. The sensor network database view (SNDV) is temporarily created and maintained at the sink node. The basic idea in DADMA is to aggregate data coming from a group of sensors or excludes some sensors from the data gathering tree. These operations are carried out according to two simple rules. First, a user can retrieve a subset of data fields available in an SNDV and aggregate data by using the following aggregate m function:

$$f_a(x) = x \ div \ m \quad (3)$$

Moreover, sensor nodes can be excluded from a query by a dilute m function as follows:

$$f_d(x) = \left(x/r\right) \bmod \left(m/r\right) \quad (4)$$

Where x is the grid location of a node with respect to one of the axes, r is the resolution in meters, and m is the aggregation (or dilution) factor.

Quantile Digest: In quantile-digest [59] a data structure is there for representing sensor readings with an arbitrary degree of approximation. The data compression algorithm adapts its behavior to the data distribution by automatically grouping the sensed data into variable size buckets of almost equal weight.

Distributed Source Coding (DSC): DSC allows joint coding of correlated data from multiple sources without explicit communication [60]. This is possible as long as the individual source rates satisfy certain constraints about conditional entropies. These techniques require that the correlation structure is available a priori at the independent encoders. These techniques are based on the Slepian-Wolf theorem [61].

Feed back control in Data Aggregation:

In AIDA [38] a dynamic feedback scheme has been proposed which is a combination of on-demand and fixed aggregation, where the degree of aggregation threshold is modified dynamically. This scheme tunes the degree of aggregation threshold and the sending rate to optimize the aggregation performance.

In [53] it define a strategy to tune the degree of data aggregation while maintaining specified latency bounds on data delivery and minimizing the energy consumption. They consider time-constrained reference scenarios dealing with real-time applications that impose specific time constraints on the delivery of sensor measurements. Data is grouped into different classes associated with different bounds on the delivery time. The aim is to guarantee the delivery of all data at the minimum energy cost while satisfying all time constraints. The data aggregation degree is adapted accordingly to meet these requirements. If the total communication load exceeds system capacity, the amount of data has to be reduced to increase the data aggregation degree, whereas the data aggregation degree may be relaxed for low traffic. This solution is interesting for two reasons. First, the control of the data aggregation is based on physical measurements of

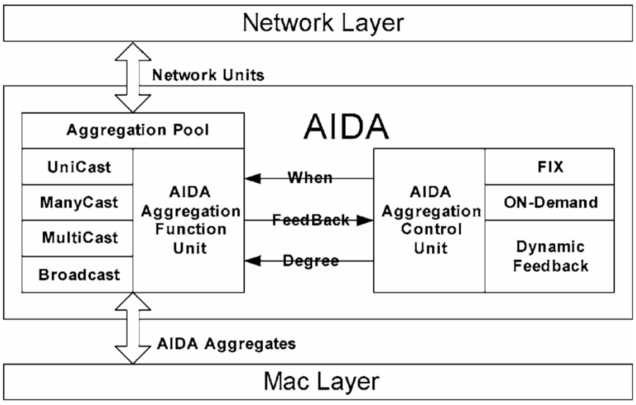

Fig.5. AIDA components [38]

the network conditions, thus making the mechanism self adaptable to the actual network dynamics. Second, it aims at satisfying time constraints that, in general, are rarely considered by wireless sensor network algorithms.

Maximum Lifetime Data Aggregation (MLDA): In [54] the goal of the MLDA problem was to obtain a data-gathering schedule with maximum lifetime where sensors aggregate incoming data packets.

Timing Strategies: Moreover, in-network aggregation techniques may require some form of synchronization among nodes. In particular, the best strategy at a given node is not always to send data as soon as it is available. Waiting for information from neighboring nodes may lead to better data aggregation opportunities and, in turn, improved performance. Timing strategies are required especially in the case of monitoring applications where sensor nodes need to periodically report their readings to the sink. These strategies usually involve data gathering trees rooted at the sink. The main timing strategies proposed so far are summarized below [16]-[18]:

- *Periodic simple aggregation:* It requires each node to wait for a predefined period of time, aggregate all data items received, and then send out a packet with the result of the aggregation.
- *Periodic per-hop aggregation:* It is quite similar to the previous approach, the only difference being that the aggregated data is transmitted as soon as the node hears from all of its children. This requires that each node knows the number of its children. In addition, a timeout is used in case some children's packets are lost.
- *Periodic per-hop adjusted aggregation:* It adjusts the timeout of a node, after which it sends the aggregated data, depending on the node's position in the gathering tree. Note that the choice of the timing strategy strongly affects the design of the routing protocol.

Comparisons of Periodic aggregation with other types of aggregation [57] are shown in Table II.

VII. Challenges of Data Aggregation and Some Existing Solutions

The main challenges of Data aggregation and some existing solutions are summarized below:

Quality of Data Aggregation Process: To ensure quality of data aggregation process is one of the main challenges of data aggregation. In [47] an information model for sensed data is first formulated. A new metric for evaluating data aggregation process, Data Aggregation Quality (DAQ), is formally derived. DAQ is used to evaluate data fusion quality implied by an underlying continuous data gathering protocol, without assuming any knowledge of the value or its statistical distribution of the sensing data.

Security: Providing hierarchical data aggregation without losing security guarantee is an interesting and challenging problem in sensor networks. Hop-by-hop data aggregation is a very important technique for reducing the

Table II: Different Approaches involved in data aggregation

Approach	Advantages	Limitations
Periodic data aggregation	Minimal control overhead and does not require clock synchronization among sensors.	This approach needs to be generalized for non-periodic data aggregation scenarios.
Real-time event Monitoring applications	Iterative numerical optimization algorithm minimizes the energy dissipation of sensors .About 20 percent to 90 percent energy savings are obtained compared to a classic radio shut down technique	There is no guarantee on the convergence speed of the iterative algorithm.
Applications that involve many to one communications such as detection in cluster based networks	In this approach study of capacity and energy consumption helps the system designer to choose particular network architecture depending on the capacity and energy constraints.	Range is unrealistic and does not consider wireless channel fading.

communication overhead and energy expenditure of sensor nodes during the process of data collection in a sensor network. However, because individual sensor readings are lost in the per-hop aggregation process, compromised nodes in the network may forge false values as the aggregation results of other nodes, tricking the base station into accepting spurious aggregation results.

In SDAP [48], a Secure Hop-by-hop Data Aggregation Protocol for sensor networks has been proposed. The design of SDAP is based on the principles of divide-and conquer and commit-and-attest. First, SDAP uses a novel probabilistic grouping technique to dynamically partition the nodes in a tree topology into multiple logical groups (sub trees) of similar sizes. A commitment-based hop-by-hop aggregation is performed in each group to generate a group aggregate. The base station then identifies the suspicious groups based on the set of group aggregates. Finally, each group under suspect participates in an attestation process to prove the correctness of its group aggregate. The aggregate by the base station is calculated over all the group aggregates that are either normal or have passed the attestation procedure.

In WSN several data inputs at an aggregator node are merged to produce compact small output data, thus reducing the network's transmission overhead and preserving the scarce-battery resources of sensors. Unfortunately if not secured against attackers, data aggregation can produce erroneous results, which can induce the network operator to take wrong decisions. In [50] it presents a secure aggregation protocol for cluster-based WSN, which is adapted to a large variety of aggregation functions, and which does not rely on trusted aggregator nodes. The inherent transmission overhead is even smaller than in other secure aggregation protocols.

Energy Constraint sensor node: The limited energy of sensors is a challenge for performing data aggregation. Since Wireless Sensor Network is composed of hundreds or thousands of sensor nodes equipped with non rechargeable and non-replaceable batteries. For such sensors, transmitting is much more energy consuming than computing, so the amount of transmitted data on the network must be kept as low as possible, in order to extend the lifetime of the network. Some approaches [20], [27], [28] are there to facilitate this.

Limited storage and Processing Capabilities: Limited storage and processing capability of sensor nodes is a challenge for performing data aggregation. But rapid advances in processor, memory, and radio technology have enabled [1]-[4] the sensor nodes to some extend to support the data aggregation process.

Link failure & Synchronization: Link failure is another important challenge for data aggregation especially for Tree based approaches. More precisely, when a packet is lost at a given level of the tree, the data coming from the related sub tree are lost as well producing erroneous result. As data is propagated towards the sink, multiple levels of data fusion are likely. The data fusion at various levels should be synchronized in order to fuse data effectively. Information from as many as sensors possible to be fused increases the accuracy of the aggregated report. In [62] it proposes a methodology by which the various levels of fusion are synchronized to ensure that the aggregated report has a desired trade-off between creditability and latency regardless of the structure of topology.

VIII. Conclusion

In this paper we have presented a detailed review of in-network aggregation techniques for wireless sensor networks. One of the main design aspects for sensor network architectures is to extend network lifetime by achieving energy efficiency. Data Aggregation techniques plays an important role for achieving energy efficiency as they aim to reduce the number of transmissions required for data collection, which, in turn, reduces energy consumption. Several approaches are presented regarding suitable routing protocols, effective aggregation functions and efficient ways of representing the data. Many solutions are proposed in the tree-based and cluster-based categories, but very few studies use the multipath and hybrid approaches. This leaves room for further work in this area. Most existing research focuses on networking issues such as routing, often considering only very simple approaches to aggregate data, but much work still remains to be done to provide cross-layer solutions, accounting for application, data representation etc.Few studies provide a deeper analysis of the aggregation functions. Previous work mostly takes spatial correlation and temporal correlation of data into account, but semantic correlation is not sufficiently well studied. So in future this can be a promising research area.

Acknowledgement

Authors are grateful to Department of Science and Technology (DST) for sanctioning a research Project under Fast Track Young Scientist scheme reference no.: SERC/ET-0213/2011 under which this paper has been completed.

References

[1] Chandrakasan, R. Amirtharajah, S.-H. Cho, J. Goodman, G. Konduri, J. Kulik, W. Rabiner, and A. Wang, "Design Considerations for Distributed Microsensor Networks," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 279-286, 1999.

[2] M. Dong, K. Yung, and W. Kaiser,"Low Power Signal Processing Architectures for Network Microsensors," *Proc. Int'l Symp. Low Power Electronics and Design*, pp. 173-177, 1997.

[3] Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks,"*Proc. ACM/IEEE Mobicom, Aug. 1999.*

[4] J. Rabaey, J. Ammer, J.L. da Silva Jr and D. Patel. "Pico Radio: Ad-hoc Wireless Networking of Ubiquitous Low-Energy Sensor/Monitor Nodes". *Proc. IEEE Computer Society Annual Workshop on VLSI*, 2000.

[5] C. Intanagonwiwat *et al.*, "Directed Diffusion for Wireless Sensor Networking," *IEEE/ACM Trans. Net.*, vol. 11,no. 1, Feb. 2002, pp. 2–16.

[6] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data Gathering Algorithms in Sensor Networks using Energy Metrics," *IEEE Trans. Parallel Distrib*. Sys. vol. 13, no. 9, Sept. 2002, pp. 924–35.
[7] Y. Xu, J. Heidemann, and D. Estrin, "Geographic-Informed Energy Conservation for Ad Hoc Routing," *ACM/SIGMOBILE MobiCom 2001*, Rome, Italy, July 2001.
[8] E. Fasolo, M. Rossi, J. Widmer, M.Zorzi "In-network Aggregation Techniques for Wireless Sensor Network: A Survey" *IEEE Wireless Communications*, vol 14, no 2, April 2007,pp. 70-87.
[9] W. Heinzelman, A. Chandrakasan, and H. balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks",*Proc.IEEE. 33 Hawaii International Conference on System Sciences*, January 2000.
[10] I. Solis and K. Obraczka, "Isolines: Energy-Efficient Mapping in Sensor Networks," *IEEE ISCC '05*, Cartagena,Spain, June 2005.
[11] V. Erramilli, I. Matta, and A. Bestavros, "On the Interaction Between Data Aggregation and Topology Control in Wireless Sensor Networks," *IEEE SECON '04*, Santa Clara, CA, Oct. 2004.
[12] M. Ding, X. Cheng, and G. Xue, "Aggregation Tree Construction in Sensor Networks," *IEEE VTC '03*, Orlando,FL, Oct. 2003,pp. 2168-2172.
[13] H. Luo *et al.*,"Energy Efficient Routing with Adaptive Data Fusion in Sensor Networks," *3rd ACM/SIGMOBILE Wksp. Foundations of Mobile Comp.*, Cologne, Germany, Aug. 2005.
[14] H. Albert, R. Kravets, and I. Gupta, "Building Trees Based On Aggregation Efficiency in Sensor Networks," *Med-Hoc-Net 2006, Lipari*, Italy, June 2006.
[15] S. Madden *et al.*, "TAG: a Tiny AGgregation Service for Ad Hoc Sensor Networks," *OSDI 2002*, Boston, MA, Dec. 2002.
[16] I. Solis and K. Obraczka, "The Impact of Timing in Data Aggregation for Sensor Networks," *IEEE ICC 2004*,Paris, France, June 2004.
[17] F. Hu, C. Xiaojun, and C. May, "Optimized Scheduling for Data Aggregation in Wireless Sensor Networks," *IEEE ITCC '05*, Las Vegas, NV, Apr. 2005.
[18] X. Jianbo, Z. Siliang, and Q. Fengjiao, "A New In-network Data Aggregation Technology of Wireless Sensor Networks," *IEEE SKG '06*, Guilin, China, Nov. 2006.
[19] K. Dasgupta, K. Kalpakis, P. Namjoshi, "An Efficient Clustering-based Heuristic for Data Gathering and Aggregation in Sensor Networks," *IEEE WCNC '03*, New Orleans, LA, Mar. 2003.
[20] J. N. Al-Karaki, R. Ul-Mustafa, and A. E. Kamal, "Data Aggregation in Wireless Sensor Networks: Exact and Approximate Algorithms," *IEEE HPSR 2004*, Phoenix, AZ, Apr. 2004.
[21] G. Hartl and B. Li, "infer: A Bayesian Approach towards Energy Efficient Data Collection in Dense Sensor Networks," *IEEE ICDCS 2005*, Columbia, OH, June 2005, pp. 371-380.
[22] H. O. Tan and I. Korpeoglu, "Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks," *ACM SIGMOD Record*, vol. 32, no. 4, Dec.2003, pp. 66–71.
[23] H. Cheng, Q. Liu, and X. Jia, "Heuristic Algorithms for Real-Time Data Aggregation in Wireless Sensor Networks, *"ACM IWCCC '06*, Vancouver, BC, Canada, July 2006.
[24] Y. Hu, N. Yu, and X. Jia, "Energy Efficient Real-Time Data Aggregation in Wireless Sensor Networks," *ACMIWCCC '06*, Vancouver, BC, Canada, July 2006.
[25] J. Choi *et al.*, "Aggregation Time Control Algorithm for Time constrained Data Delivery in Wireless Sensor Networks," *IEEE VTC 2006-Spring*, Melbourne, Australia, May 2006.
[26] H. Gupta *et al.*, "Efficient Gathering of Correlated Data in Sensor Networks," *ACMMobiHoc 2005*,Urbana-Champaign, IL, May 2005.
[27] Hyun-sook Kim and Ki-jun Han , "A Power Efficient Routing Protocol Based on Balanced Tree in wireless Sensor Networks" Proc. IEEE International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05).
[28] Hu¨ seyin Ozgu¨r Tan, Ibrahim Korpeoglu, 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,pp. 489-500.
[29] B. Zhou *et al.*, "A Hierarchical Scheme for Data Aggregation in Sensor Network," IEEE ICON '04, Singapore, Nov. 2004.
[30] C. Intanagonwiwat *et al.*, "Impact of Network Density on Data Aggregation in Wireless Sensor Networks," *IEEE ICDCS '02*, Vienna, Austria, July 2002.
[31] S.Lindsey and C.Raghavendra, "PEGASIS: Power Efficient Gathering in Sensor Information Systems",Proc. IEEE. Aerospace Conference, March 2002,pp.1-6
[32] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan,"An application-Specific Protocol Architecture for Wireless Microsensor Networks,"*IEEE Trans. Wireless Commun.*, vol. 1, no. 4, Oct. 2002, pp. 660–70.
[33] Y. Yao and J. Gehrke, "Query Processing for Sensor Networks," *ACMCIDR 2003*, Asilomar, CA, Jan. 2003.
[34] A. Mahimkar and T. S. Rappaport, "SecureDAV: A Secure Data Aggregation and Verification Protocol for Sensor Networks," *IEEE GLOBECOM 2004*, Dallas, TX, Nov. 2004.
[35] Y. Yao and J. Gehrke,"The Cougar Approach to In Network Query Processing in Sensor Networks," *ACM SIGMOD Record*, vol. 31, no. 3, Sept. 2002, pp. 9–18.
[36] S. Nath et al., "Synopsis Diffusion for Robust Aggregation in Sensor Networks," *ACM SenSys 2004*,Baltimore, MD,Nov. 2004.
[37] A. Manjhi, S. Nath, and P. B. Gibbons, "Tributaries and Deltas: Efficient and Robust Aggregation in Sensor Network Stream," *ACM SIGMOD 2005*, Baltimore, MD, June 2005.
[38] T. He et al., "AIDA: Adaptive Application-Independent Data Aggregation in Wireless Sensor Networks," *ACM Trans. Embedded Computing Systems*, vol. 3, no. 2, May 2004, pp. 426–57.
[39] U. Ramachandran et al., "Dynamic Data Fusion for Future Sensor Networks" *ACM Trans. Sensor Networks*, vol. 2, no. 3, Aug. 2006, pp. 404–43.
[40] M. Wenz , H. Worn, "Event-based Production Rules for Data Aggregation in Wireless Sensor Networks," *IEEE MFI '06, Heidelberg*, Germany, Sep. 2006.
[41] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, Oct. 2002, pp. 660–70.
[42] K. Akkaya and M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks," *Elsevier Ad Hoc NetworkJ.*, vol. 3, no. 3, May 2005, pp. 325–49.
[43] J. N. Al-Karaki and A. E. Kamal, "Routing Techniques in Wireless Sensor Networks: a Survey," *IEEE Wireless Commun.*, vol. 11, no. 6, Dec. 2004, pp. 6–28.
[44] Holger Karl, Marc Lobbers, Tim Nieberg, "A Data Aggregation Framework for Wireless Sensor Networks," *In Proc. Dutch Technology Foundation ProRISC Workshop on Circuits, Systems and Signal Processing*, Nov. 2003.
[45] Nan Li, Shangru Li, Xiaozhou Fang "Adaptive Data Aggregation Mechanism Based On LEACH Protocol" *Proceedings of AIAI2010.*
[46] Jiao Zhang , Fengyuan Ren, Tao He , Chuang Lin "Attribute-aware Data Aggregation Using Dynamic Routing in Wireless Sensor Networks", *proc IEEE*, pp. 1-9.
[47] Tri Pham, Eun Jik Kim, Melody Moh "On Data Aggregation Quality and Energy Efficiency of Wireless Sensor Network Protocols - Extended Summary",Proc. *IEEE, International Conference on Broadband Networks (BROADNETS'04).*
[48] Y. Yang, X. Wang, S. Zhu, G. Cao "SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks" *ACM Transactions on Information and Systems Security*, Vol. 11, No. 4, July 2008.
[49] K. Kalpakis, S. Tang "Collaborative data gathering in wireless sensor networks using measurement co occurrence" *Elsevier, Computer Communications 31* ,2008 ,pp.1979–1992.
[50] Chakib Bekara , Maryline Laurent-Maknavicius, "A Secure Aggregation Protocol for Cluster-Based Wireless Sensor Networks With no Requirements for Trusted Aggregator Nodes", *Proc. IEEE, International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2007).*
[51] E. Cohen and H. Kaplan, "Spatially-Decaying Aggregation Over a Network: Model and Algorithms," *ACM SIGMOD'04*, Paris, France, June 2004.
[52] A. Sharaf et al., "Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks," *VLDB J.*, vol. 13, no. 4, Dec. 2004, pp. 384–403.

[53] T. Abdelzaher, T. He, and J. Stankovic, "Feedback Control of Data Aggregation in Sensor Networks," *IEEE CDC '04*, Atlantis, Paradise Island, Bahamas, Dec. 2004.
[54] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Efficient Algorithms for Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks," *Computer Networks*, vol. 42,no. 6, Aug. 2003, pp. 697–716.
[55] A.Ray, D.De "Energy Efficient Clustering Hierarchy Protocol for Wireless Sensor Network" Proc. *IEEE International Conference on Communication and Industrial Application (ICCIA)* 2011, pp.1-4.
[56] A.Ray, D.De "Energy Efficient Cluster Head Selection in Wireless Sensor Network" Proc. *IEEE Recent Advances In Information technology(RAIT)*,ISM Dhanbad- March 2012.
[57] R.Rajagopalan, P.K.Varshney "Data Aggregation Techniques in Sensor Network: A Survey", *IEEE Communications Surveys & Tutorials,4th Quarter 2006,* Vol. 8, No. 4, pp.48-63.
[58] E. Cayirci, "Data Aggregation and Dilution by Modulus Addressing in Wireless Sensor Networks," *IEEE Commun. Lett.*, vol. 7, no. 8, Aug. 2003, pp. 355–57.
[59] N. Shrivastava *et al.*, "Medians and Beyond: New Aggregation Techniques for Sensor Networks," *ACM SenSys '04*, Baltimore, MD, Nov. 2004.
[60] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed Source Coding for Sensor Networks," *IEEE Signal Processing*, vol. 21, no. 5, Sept. 2004, pp. 80–94.
[61] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.
[62] Wei Yuan, Srikanth V. Krishnamurthy, and Satish K. Tripathi "Synchronization of Multiple Levels of Data Fusion in Wireless Sensor Networks" Proc *IEEE GLOBECOM 2003*, pp.221-225.

Anindita Ray received the B.Sc. degree (with honours) in Physics and the MCA degree from the University of North Bengal, Darjeeling, India, in 1999 and 2003, respectively. She is M.Tech. in Software Engineering in 2012 from West Bengal University of Technology, Kolkata, India. She has seven years teaching experience, five and half years as a Lecturer and one and half year as an Assistant Professor in the Department of Computer Application, B. P. Poddar Institute of Management and Technology under West Bengal University of Technology, Kolkata, India. Her research interest includes wireless sensor network, mobile computing and soft computing. Her email id is aninrc@gmail.com

Dr. Debashis De received M.Tech degree in Radio Physics & Electronics in 2002. He obtained his Ph.D (Engineering) from Jadavpur University in 2005. He worked as R& D Engineer of Telektronics. Presently he is an Associate Professor in the Department of Computer Science and Engineering of West Bengal University of Technology, India and Adjunct Research Fellow of University of Western Australia, Australia. He was awarded the prestigious Boyscast Fellowship by department of Science and Technology, Govt. of India to work at Herriot-Watt University, Scotland, UK. He is also awarded Endeavour Fellowship Award during 2008-2009 by DEST Australia to work in the University of Western Australia. He received Young Scientist award both in 2005 at New Delhi and in 2011 at Istanbul by International Union of Radio Science, H.Q.,Belgium. His research Interest includes Location Management and Power Consumption Control in Mobile network and low power Nano device design for mobile application and disaster management. His email id: dr.debashis.de@gmail.com