**Data Fusion in Wireless Sensor Network-A Survey**

**Ishita Chakraborty**, **Prodipto Das**

1 Department of Computer Science, Assam University, Silchar, India  
2 Department of Computer Science, Assam University, Silchar, India  
*Corresponding Author: ishitachakraborty18@gmail.com*

Received: 15/Nov/2017, Revised: 26/Nov/2017, Accepted: 22/Dec/2017, Published: 31/Dec/2017

**Abstract**—Wireless Sensor Network is a network with a number of sensor nodes deployed in a geographical area. These nodes are deployed to various kinds of environmental conditions like humidity, noise, temperature etc. The sensors are used for collecting data related to varying environmental conditions where they are deployed in. The primary constraint in a WSN is energy. Since the sensor nodes are battery operated, collecting and transmitting data will cost a lot of energy resources. Therefore, to maximize the lifetime of sensor networks, aggressive energy optimization techniques have to be used for ensuring that energy is conserved for the sensor nodes. This paper discusses about wireless sensor network, its architecture, data aggregation or fusion related algorithms.

Keywords— WSN, data aggregation, data fusion, sensor network, IoT

**I. INTRODUCTION**

A Wireless Sensor Network is a network which comprises of a number of small autonomous sensor nodes called motes. These sensor nodes monitor a physical environment situation such as temperature, motion or pressure etc. Every node in a wireless sensor network is basically a computer with attached sensors that can process and exchange sensed data[1]. They can also communicate wirelessly among each other. Sensors attached to this node allow them to sense various phenomena within the surroundings. The sensor nodes are of limited energy. So, a fundamental issue in WSN is the way to process the collected data in an energy efficient manner. Data aggregation or data fusion can be defined as the systematic collection of sensed data from multiple sensors that is to be eventually transmitted to the base station for processing[2]. The sensor nodes are battery operated so there is energy constraint in the sensor network. As a result, it is inefficient for all the sensors to transmit the data directly to the base station. The data accumulated from the neighbouring sensors are highly redundant and the amount of data which is sensed is also enormous in size. So, it is bound to energy as well as time constraint to process and transmits such a huge amount of data. Therefore, we need some techniques for combining such large amount of sensed data into high quality information at the sensors or the intermediate nodes, which can reduce the number of packets that is to be transmitted to the base station and thus conserving energy and bandwidth. This can be achieved by data aggregation. Data aggregation is considered as an important paradigm for wireless routing in sensor networks. The main task is to combine the huge amount of data coming from different sources thereby eliminating redundancy and thus minimizing the number of transmissions which in turn saves energy[3]. There are numerous applications of wireless sensor networks which typically involves some kind of monitoring, tracking, or controlling of physically occurring events. Area monitoring, environmental data collection, landfill ground well level monitoring and pump counter, security monitoring, vehicle detection, agriculture and green house monitoring etc.[1].

**II. RELATED WORK**

There has been a lot of development proceeding in the field of data aggregation techniques of which cluster based and tree based data aggregation techniques are used very commonly. In cluster based data aggregation, the entire network is divided into several cluster regions and cluster head node is elected in each and every cluster regions, which is responsible for receiving and sending data to sink node, such as LEACH, TEEN, etc [4]. Haowen Chan et al in [5] proposed an algorithm based on secure information aggregation (SIA) in which the maximum, minimum, average, and sum functions, are used to finish data fusion and transmit the results of integration to the base station. Chen Hui-fang in [6] proposed an adaptive data aggregation (ADA) scheme for the clustered WSNs. Tree-based fusion method aims at collecting data from the distributed sensor nodes hierarchically through the reverse tree where the source node is a leaf node of the data fusion.
tree [4]. In [7] B. Krishnamchari et al proposed energy savings and the delay tradeoffs involved in data aggregation and also investigate the computational complexity of optimal data aggregation in sensor. In [8], Wen-Hwa Liao et al proposed an ant colony algorithm for data aggregation in wireless sensor networks. In this mechanism, artificial ants are assigned to the source nodes to establish low latency path between the source and the sink nodes. Some other techniques of data aggregation include the directed diffusion technique, in-network aggregation, fuzzy-based data fusion etc. [2], [9], [10].

III. ARCHITECTURE OF WSN

The architecture of a sensor node showed in Figure 1 consists of four major parts- Energy supply, sensor, analog to digital converter (ADC), processor and memory, and transceiver. Transceivers are used for sending and receiving data simultaneously. The power unit or the energy supply unit is required for powering the sensor nodes. The sensor circuitry converts the physical quantities into an electric signal. The ADC changes the analog signal generated by the sensor into a digital signal and sends it to the processor. The processor then executes operations on the received digital signal, and can store it into memory [3]. The architecture of wireless sensor network includes a hardware platform and an operating system which is designed specifically to meet with the needs of wireless sensor networks. TinyOS is a component-based operating system which is designed to run in resource constrained wireless devices. TinyOS has been designed in such a manner that it can run on a generalised architecture using one CPU which is shared between applications and protocol processing [1].

IV. TYPES OF DATA FUSION

A. Low-level fusion

In this type of fusion, the raw data are provided as inputs and they are combined into new data that are more accurate than the individual inputs [11].

B. Medium-level fusion

Here the features and the attributes of an entity are fused to obtain a feature map that may be used for other tasks. It is also known as feature/attribute level fusion.

C. High-level fusion

It is known as symbol or decision level fusion. It takes decisions or symbolic representations as input and combines them to obtain a more confident and/or a global decision. An example of high-level fusion is the Bayesian approach for binary event detection proposed by Krishnamchari in [7] that detects and corrects measurement faults.

D. Multilevel fusion

Multilevel fusion process consists of data from different abstraction levels where both the input and the output of fusion can be of any level. Data fusion is performed to eliminate the redundant data transmission, in turn enhancing the lifetime of the network. It deals with the aggregation of data from multiple sensors nodes at some intermediately node and transmitting these aggregated data to the sink [2]. The simplest form of data fusion is duplicate suppression [3], where if two or more sensor nodes send same data packets to a particular node where fusion takes place, then it forwards only a single copy of data out of the two by reducing the traffic on the network and conserves the energy of the sensors. If the cluster head (fusion node) is attacked by some malicious attacker, then it compromises with the data because the cluster head itself sends the fused data to the entire network. The performance measures of data fusion algorithms are network lifetime, latency, Communication Overhead and data accuracy. Energy efficient data fusion in wireless sensor networks are necessary because, the sensor nodes are battery operated, and it is important to keep track of the energy issues [12]. The loss of battery or energy may lead to failure of the entire network [13], [14]. Few of the energy efficient data fusion techniques have already been listed below [3]

i. Witness based data fusion- In this type of data fusion approach, the fusion node doesn’t forward the data directly to the base station. Rather it computes the Message Authentication Code (MAC) of the data. Only after receiving the MAC of the data, it forwards it to the base station.

ii. Dynamic Data Fusion- Here fusion of data is performed in a framework which is known as Dfuse and it requires advanced fusion methods.

iii. Multi-sensor data fusion- It uses heterogeneous sensor nodes which can detect parameters like temperature, humidity, pressure, carbon monoxide etc. These sensor signals can be fused by using fuzzy rule based methods [15].
iv. Single Mobile Agent-based autonomic data fusion - This technique is effective mainly for small scale WSN as it uses only one mobile agent for performing data fusion.

v. Multiple Mobile Agent-based autonomic data fusion - Large scale sensor networks support multi mobile agent-based autonomic fusion technique. In this method, multiple mobile agents work in parallel to perform fusion.

vi. Mobile Agent-based clustering data fusion - Here, two cluster head models is used to control the cluster size. All the sensor nodes are divided into clusters. Mobile agents are used in between the cluster heads for performing data fusion [16].

V. DATA AGGREGATION PROTOCOL

The architecture of a sensor network is a vital factor for determining which data aggregation protocol will be best suited. On the basis of network architecture, the following aggregation schemes are proposed

A. Flat Network

In a sensor network with flat architecture, every sensor node is treated equal and all of them participate in routing the data packets to the sink. Data aggregation is performed by data centric routing where the sink sends a query message to all the sensor nodes in the network via flooding and the sensors with the matching data responds back [2]. There are mainly three approaches to data fusion in flat network

1) Push Diffusion: This scheme defines the sources as active participants that initiate the diffusion while the sinks respond to the sources. Whenever they detect an event, the sources flood the data to the sink. An example of push diffusion scheme can be the sensor protocol for information via negotiation (SPIN) protocol [17]. Negotiation and resource adaption are two main features of push diffusion. In SPIN protocol topological changes are localized. The main problem with SPIN is its inability to guarantee delivery of data.

2) Two phase pull diffusion: An energy efficient data aggregation protocol called directed diffusion has been developed by Intanagonwiwat et al. [18]. It is a representative approach of two way pull diffusion. Directed diffusion is a data centric routing scheme which is based on the data received at the sensors. Figure 2 shows how interest propagation takes place in directed diffusion. If the attributes generated by the source node matches the interest, then a gradient is set up which identifies the data being generated by the sensor node. An interest message is initially broadcasted by the sink to the network. The gradient specifies the data rate and the direction in which to send the data. A data cache is maintained in each sensor node which keeps track of recently seen data items. After receiving low data rate events, the sink reinforces one particular neighbour in order to attract higher quality data. Thus, directed diffusion is achieved by using data driven local rules. Some of the factors which influence performance of data aggregation in directed diffusion protocol are position of source and destination nodes and network topology. Krishnamachari et al. [19] explained the various impacts of source-destination placement and the energy costs associated with it.

![Figure 2. Interest propagation in directed diffusion.](image)

3) One phase pull Diffusion: Too many source and sink in two phase pull diffusion, results in large overhead. In [20] the authors have proposed a one phase pull diffusion technique in which it skips the flooding process of directed diffusion. Here the sink nodes send interest messages that travel through the network establishing gradients. However, no exploratory data is transmitted by the sources. The source transmits data only to the sink with lowest latency gradient. Simulation results on comparing push diffusion with one phase pull diffusion shows that the latter outperforms the former when the source event rate is high but when there is high sink interest rate, push diffusion performs better.

B. Hierarchical Network

Flat network may cause excessive communication and computation burden at the sink node resulting in a faster depletion of its battery power and if the sink node dies, it breaks down the functionality of the network. In hierarchical data aggregation, the data fusion occurs at special nodes, which reduces the number of messages transmitted to the sink. It also improves the energy efficiency of the network. The following are the approaches to hierarchical data aggregations are

1) Cluster based data aggregation: As sensor network are energy constrained, it is inefficient for the sensors to directly send data to the sink. In such scenarios, the sensors can send their data to a local fusion node or cluster head, which aggregates data from all the sensors in its cluster and transmits the aggregated data to the
sink. Significant amount of energy can be saved for the energy constrained sensors. Figure 3 shows a cluster based sensor network organization. Some of the proposed hierarchical cluster based data aggregation protocols are Low Energy Adaptive Clustering Hierarchy (LEACH), Hybrid Energy Efficient Distributed Clustering Approach (HEED) and clustered diffusion with dynamic data aggregation (CLUDDA).

In the LEACH protocol, the sensor nodes organize themselves into clusters for performing data fusion. The nodes belonging to the particular cluster send their data to the respective cluster head. The cluster head in each cluster transmits the fused data from multiple sensors in its cluster to the sink node. This reduces the amount of data transmitted to the sink. The data aggregation in LEACH is performed periodically at the cluster heads. A hybrid approach has been proposed in [22] in which clustering is combined with diffusion techniques. This new aggregation scheme is called clustered diffusion with dynamic data aggregation (CLUDDA). In CLUDDA, directed diffusion with clustering is performed in the initial phase of query generation. The Clustering ensures only the cluster heads to participate in inter-network communication for transmitting the interest messages throughout the network. This ensures energy optimization in the network as the other nodes remain inactive during the query generation phase. An interesting feature of CLUDDA algorithm is that a query cache is maintained at the cluster heads which contains all the data components that were aggregated to generate the final data. It also contains the neighbouring node addresses from which data was generated.

3) Tree based data aggregation: Here the sensor nodes are organised in a tree structure and the data aggregation is performed at intermediate tree nodes. A prime aspect of tree based aggregation scheme is to constructing an energy efficient data aggregation tree. Ding et al. [23] proposed a tree based data aggregation algorithm called energy aware distributed heuristic aggregation tree (EADAT) to construct and maintain a data aggregation tree in sensor networks. In this algorithm, the sink initiates by broadcasting a control message to the entire network and acts as the root node in the aggregation tree. The control message contains the five fields: ID, parent, power, status and hpcnt which indicates the sensor ID, its parent node, residual power, the status (leaf, non-leaf node or undefined state) and the number of hops from the sink node. On simulating the above protocol on a sensor field of 160m x 160m, it proved to conserve more energy in comparison with routing methods without aggregation.

4) Grid based data aggregation: Vaidhyanathan et al. [24] have proposed two data aggregation schemes in which they divide the region into several grids. They have proposed two methodologies, namely, grid based data aggregation and in-network data aggregation. In grid based aggregation scheme, a set of sensor nodes are assigned in a particular region of the sensor network as data aggregator. Sensor nodes in the fixed region will transmit their data directly to the data aggregator in the grid. Hence, the sensors within a fixed grid do not communicate with each other and the data aggregator is fixed in each grid. In contrast, in-network data aggregation enables communication of the sensors with its neighbours within the same grid. This means that any node can play the role of data aggregator in the grid and the sensor transmits its signal strength to its neighbouring sensors. If the neighbour sensor has higher signal strength, the sender stops transmitting. After receiving all the data packets, the node with
highest signal strength is considered as the data aggregator. Figure 4 and Figure 5 shows how grid based data aggregation and in-network data aggregation takes place.

![Figure 4. Grid based data aggregation scheme.](image)

![Figure 5. In-network data aggregation scheme.](image)

The arrows in the grid based network indicated in Figure 4 are the transmission of data from sensors to the grid aggregator. The numbers and arrows indicate in figure 5 represents the signal strengths detected by the sensors and the exchange of signal strengths between neighbouring nodes respectively.

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Hierarchical networks</th>
<th>Flat networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Data aggregation is performed by cluster heads or a leader node.</td>
<td>Data aggregation is performed by different nodes along the multi-hop path.</td>
</tr>
<tr>
<td>2.</td>
<td>Overhead is involved in cluster or chain formation throughout the network.</td>
<td>Data aggregation routes are formed only in regions that have data for transmission.</td>
</tr>
<tr>
<td>3.</td>
<td>Even if one cluster head fails, the network may still be operational.</td>
<td>The failure of sink node may result in the breakdown of entire network.</td>
</tr>
<tr>
<td>4.</td>
<td>Lower latency, since sensor nodes perform</td>
<td>Higher latency is involved in data transmission to the</td>
</tr>
</tbody>
</table>

| 5.    | Routing structure is simple but not necessarily optimal. | Optimal routing can be guaranteed with additional overhead. |
| 6.    | Node heterogeneity can be exploited by assigning high energy nodes as cluster heads | Does not utilize node heterogeneity for improving energy efficiency. |

The internet has significantly expanded its capabilities by facilitating connections between large numbers of heterogeneous devices, called as the Internet of Things (IoT). Many areas such as agriculture, automobile, electric power distribution etc. which were initially unaffected by the Internet will be devastated by this network extension [25]. IoT is such a huge paradigm that acts as a bridge between the diversely available and recent technologies like WSN, cloud computing and information sensing [26]. To collect and process data effectively and efficiently. IoT provides a platform for WSNs to connect to the internet and experience the power of cloud computing and data fusion. In [27], Gnawali et al proposed a state-of-the-art routing protocol for sensor network in which the nodes forward data directly to the sink. A distinguishing factor in IoT is that data is acquired from many sources. In an IoT application, these acquired data from multiple sources are periodically collected and forwarded to the gateways using a multi-hop routing protocol. These gateways then decide which data is to be further forwarded to the cloud.

![Figure 6. A cloud-based IoT network model](image)
VI. CONCLUSION

The application of Wireless Sensor Networks has attracted significant attention over the past few years. Since it is battery operated, it is important to optimise the energy of the network by grouping the nodes into clusters, which has by far been the most widely used approach for maximizing the lifetime of the sensor network. Since 70% of the energy is used in data transmission in a wireless sensor network, it is a very useful technique to optimize the data transmission which is taking place in the wireless sensor network. Data transmission can be optimized with the use of effective routing protocols and data fusion methods. In this paper, the types of data fusion techniques, various approaches of data fusion and its application in IoT have been discussed. With the use of these techniques, the load on the sensor network can be decreased, thereby increasing the lifetime of the network.

REFERENCES


Authors Profile

Ms. Ishita Chakraborty passed her Bachelor of Information Technology from Sathyabama University, Tamil Nadu, India, in 2009 and Masters in Information Technology from NERIST, Arunachal Pradesh, India, in year 2011. From 2011 she is working as Assistant Professor in Department of Computer Engineering at Royal Global University, Guwahati, India. Her main research work focuses on routing techniques, Sensor networks, IoT and Network Security. She has 7 years of teaching experience.