A Comprehensive Survey of Data Processing Approaches

Shivangi Katiyar
Department of Computer Science and Engineering,
M.M.University, Solan, Himachal Pradesh, India.

Devendra Prasad
Department of Computer Science and Engineering,
M.M.University, Mullana, Ambala, Haryana, India.

ABSTRACT
In Wireless Sensor Networks (WSNs), energy of the sensor nodes are the most concerning factor because each sensor is equipped with limited amount of energy which is used to perform lots of work. Sensors have capabilities of sensing, analyzing, processing and communication of the data. In WSNs, the most responsible factor behind energy consumption of the sensor nodes is Data Processing which mainly involves data sensing, data aggregation and dissemination. Sensor nodes needs to process data in such a way that it can generate useful information on spending only affordable amount of energy. The past records shows that in early stages WSNs architecture was based on static nodes of densely deployed sensing areas and after that authors have also proposed WSNs architecture based on mobility with the help of pre-existing techniques and new ideas. A comprehensive survey of existing data processing approaches and their multilevel classification scheme is presented in this paper. The taxonomy together with the comparative tables can be used as a guideline to select a technique suitable for their intended application.

Keywords
Data dissemination, Data aggregation, Mobile Elements

1. INTRODUCTION
A wireless sensor network (WSN) consists of number of tiny sensors that are distributed over a vast field to monitor and interact with the surrounding environment. These tiny nodes have to collect data from the environment, process that data for retrieving meaningful information and share that aggregated data over a communication link. These small nodes are equipped with limited amount of battery life, communication range and storage capabilities [1]. In WSNs, a potentially large number of sensors generate a substantial amount of data which needs to design and develop techniques to efficiently process that data [2]. The concept of mobility in WSNs comes to satisfy the increasing demand of the applications of different fields and to prolong the network lifetime. Till today authors had been proposed lots of data processing techniques which can be classified into two types, i.e. static data processing and dynamic data processing. Both static and dynamic data processing approaches focuses on lower energy consumption and better performance of the sensor nodes, the difference exists is mobility factor in dynamic approaches which creates huge topological difference between these two. In static approaches, nodes remains fixed at their own position whereas in dynamic approaches there is mobility in sink node, regular nodes and mobile agent. This paper serves a comprehensive analysis of static and dynamic data processing approaches. The rest of the paper is organized as follows: In section II discuss about the Data processing and factors which basically affects data processing in WSNs. Section III deals with classification of Data processing approaches and based on analysis we provides useful information about these approaches.

2. DATA PROCESSING
Data processing has been affected by two factors i.e. number of sink present in the network and the type of approach used to process data. In WSNs the deployed sensors collect the sensed data and sends to the sink or base station for further processing, whereas network may include single sink or multiple. Early deployments of WSNs supported many-to-one paradigm; single sink node collects data from large number of sensing devices therefore communication protocols are geared towards the efficient and reliable transactions to a single receiver. Single receiver approaches failed to provide efficient solutions for more decentralized scenarios like in sensor networks [3], which inherently call for routing solutions to report to multiple receivers. Therefore recent deployment increasingly calls for many-to-many paradigm i.e. the sensed data must be delivered to multiple sinks. If talk about the data processing then generally, approaches that deal with decisions about the amount of data to be processed or communicated and how this processing is possible; can be classified as Data Processing approaches.

3. CLASSIFICATION OF DATA PROCESSING APPROACHES
The overall classification of data processing approaches is shown in Fig.1 that efficiently utilizes node’s energy, thereby prolong sensor network lifetime.

3.1 Static Data Processing approaches
In Static data processing approaches, regular nodes as well as sink node remains fixed at their own position throughout the network life time which is based on prefixed infrastructure. In this kind, regular nodes regularly transfer the sensed data to the base station/sink node. Static data processing approaches can be further classified in to two types i.e. (i) Network Architecture driven (ii) Data driven.

3.1.1 Architecture driven
The Architecture of WSNs plays an important role in the performance of Data processing techniques. Based on various applications and requirements, possible classification of Architecture driven techniques are discussed below and Table I Shows the summary of Architecture driven Approaches.
3.1.1.1 Flat architecture based
For the small scale WSNs, Flat architecture is very effective. Data processing in flat architecture uses Flooding and Gossiping. In [4], authors present a data dissemination protocol i.e. Sensor Protocols for Information via Negotiation (SPIN). This protocol is basically designed to address the drawbacks of flooding and gossiping. In [5] author present Rumor Routing, which is basically logical compromise between flooding queries and flooding event notifications.

In [6], author presents Trajectory based Forwarding (TBF) to forward packets in dense network; this is basically generalization of source based routing and Cartesian forwarding.

3.1.1.2 Cluster based
Earlier, WSNs architecture was based on flat architecture which seems whole network as a single network and do not consider any sub-network. WSNs based on this type of architecture suffered from number of problems like sink failure, energy hole problem etc. In order to address above issues, sensor nodes are partitioned into various groups called clusters [7]. In [8] author proposed an Energy-Efficient Communication Protocols for Wireless Micro-sensor Networks i.e. Low Energy Adaptive Clustering Hierarchy (LEACH) and it is considered as basic energy efficient hierarchical routing protocol. LEACH protocol is a routing protocol but can also act as aggregation protocol. In [9] author proposed, Power-Efficient Gathering in Sensor information system (PEGASIS) and can be considered as further optimization of the LEACH algorithm. In PEGASIS it is assumed that location of all nodes is known to every node and each node communicates only with its close neighbor present in cluster. In [10], authors proposed distributed clustering and data aggregation algorithm (CODA), in this method nodes are clustered according to their distance from the sink node. After collecting the data, data is aggregated using unsupervised learning, which is the Kohonen Self Organizing map (SOP).

3.1.1.3 Tree based
Tree based data processing approaches gets attracted by authors because of its execution ease and appreciated results. In [11], authors proposed routing schemes based on near optimal minimum spanning tree. In [12], authors proposed heuristics to construct and maintain an aggregation tree in sensor networks to facilitate data-centric routing. In the tree, only the non-leaf nodes are in charge of data aggregation to save the number of broadcasting messages and radio of leaf nodes are tuned of most of the time to save power and to prolong the network lifetime. In [13], converge-casting tree construction and channel allocation algorithm (CTCCAA) constructs a tree for collision free converge-casting. In this scheme time slot is compared by calculating total duration of aggregation. In [15] authors proposed a distributed algorithm for the construction of a minimum spanning tree with hierarchical clusters which is able to save energy efficiently.

3.1.1.4 Grid based
In [16] author proposed a Cycle-Based Data Aggregation Scheme for grid-based WSNs. Scheme, first construct the grid infrastructure by partitioning the whole network into a grid of cells and in each cell the node with most residual energy is chosen to be the cell head to evenly distribute the energy depletion.

![Classification of Data processing approaches in WSNs](image)

**TABLE I. Classification of Architecture Driven Approaches**

<table>
<thead>
<tr>
<th>Approach</th>
<th>Features</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| Flat network     | + All nodes play the same role and data communication is achieved in multi-hop fashion.  
                   + Data aggregation routes are formed only in region that have data for transmission.  
                   + Less overhead to organize the network and its traffic that’s why not more efforts required in processing data.  
                   + Data processing in mobile environment is also possible.                                                                     | - In Direct Transmission nodes far from Base Station (BS) suffer more energy dissipation than those close to BS.  
                   - In Minimum Transmission Energy nodes closest to BS will die first.  
                   - Failure of sink results whole network failure.  
                   - Not supportive in large scale WSNs because of higher latency and unreliable data delivery.                           |
| Applications     | It is very effective for small scale WSNs.                                |                                                                             |
| Belonging schemes| SPIN[4], Rumor Routing[5], TBF[6] etc.                                    |                                                                             |
### Cluster Network
- Better utilization of node’s battery power.
- Clustering on the basis of node’s residual energy thereby prolong network lifetime.
- Easily scalable up to thousand nodes.
- Dynamic to topology changes.
- Highly reliable and flexible.

<table>
<thead>
<tr>
<th>Application</th>
<th>For Large scale WSNs, it is best suitable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging Schemes</td>
<td>LEACH[8], PEGASIS[9] etc.</td>
</tr>
</tbody>
</table>

### Tree Network
- Simple and efficient to deliver data from several nodes to sink.
- Balanced energy consumption.
- Only Local knowledge of topology is required.

<table>
<thead>
<tr>
<th>Application</th>
<th>Best suitable for application which involves In-network data aggregation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging Schemes</td>
<td>PEDAP[11], Enhanced-DD [14], TECARP [40] etc.</td>
</tr>
</tbody>
</table>

### Grid Network
- It partitions the whole network into grid with location information which results fast and reliable data processing.
- Highly energy efficient.
- Comfortable with mobility.

<table>
<thead>
<tr>
<th>Application</th>
<th>Best suitable for application where network is location aware.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging Schemes</td>
<td>CODA[17], GBDAS [41], GBDD [42]</td>
</tr>
</tbody>
</table>

In [17] authors proposed Coordination based Data Dissemination (CODE) protocol which divides the complete sensor field into grids. Each grid has one coordinator which acts as an intermediate node. It establishes a data dissemination path so that the source can send data to the mobile sink. In [18] authors proposed an energy-efficient data-dissemination (EEDD) protocol in wireless sensor networks which additionally divides the whole sensor field into small virtual grids, every grid has a head to forward data.

#### 3.1.2 Data Driven
The aim of these approaches is to reduce the amount of total data that has to travel. The most logical way to reduce the energy consumption in WSNs is to reduce the number of radio transmissions. Various classifications of Data Driven approaches and Information are summarized in Table II.

#### 3.1.2.1 Compression based
In [19] authors proposed Pipelined In-Network COMpression (PINCO) scheme which is used in a scenario where full data is required, in such applications compression is required rather than any technique that cutout the data to reduce size. The data funneling scheme proposed by authors in [20], which achieves energy-efficient packet transmission through combining two contributing components: (i) packet aggregation and (ii) data compression. Packet aggregation scheme reduces the probability of packet collisions and compression technique reduces the data size efficiently. In [21] author presents an entropy-based compression algorithm that uses the baseline JPEG algorithm for compressing DC (direct current) coefficients. In [22] author proposed compression algorithm for WSN; instead of presenting new compression algorithm, the proposed algorithm uses combined form of pre-existing algorithm in smart way.

#### 3.1.2.2 Prediction based
Data prediction techniques usually maintain two instances of a model in the network, one residing at the sink and the other at the sensor node. The model at the sink can be used to answer queries without requiring any communication, so the original data can be easily regain within a certain degree of precision. In [23] authors investigated time-series forecasting technique for WSN based on LMS algorithm and apply on two networks i.e. on star network and cluster network. In [24] author uses neural networks for efficiently manage the power consumption in Wireless Sensor Networks (WSN). In this work a multi-layer perceptron (MLP) is used to decide on the data samples required and conduct an experiment for effective monitoring of environmental conditions to justify own. In [25] author proposed a combination of techniques i.e. wake-up receivers, node level power management and prediction-based data collection. These combinations of techniques address energy savings at the hardware and application levels.

#### 3.1.2.3 Correlation based
The correlation in WSN is of two types i.e. Spatial Correlation and Temporal Correlation. Spatial Correlation: Typical WSN applications require spatially dense sensor deployment. The key characteristic of such applications is that nearby sensor nodes monitors an environmental feature and register these values. Due to highly dense environment, sensors, observations are highly correlated. It works in variety of applications such as air pollution monitoring. Temporal Correlation: In this, sensor nodes periodically perform observations and send sensed data to the sink.
3.2 Dynamic data processing approaches

Based on mobility Dynamic Data processing Approaches are categorized in to following three types: Sink mobility based Sensor nodes mobility based and Agent mobility based. The various classifications of Dynamic Data processing Approaches are shown in Table III.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Features</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compression based</td>
<td>+ Data compression reduces the size of data packets.</td>
<td>- Needs of data packet encoding at the sensory nodes and decoding at the base station.</td>
</tr>
<tr>
<td></td>
<td>+ Retains the accuracy of collected data from sensor nodes.</td>
<td>- Data compression techniques are computationally heavy.</td>
</tr>
<tr>
<td></td>
<td>+ Due to coding-decoding concept, it is more secure than other approaches.</td>
<td>- Increased Latency.</td>
</tr>
<tr>
<td></td>
<td>+ Here data can be recompressed without decompression which reduces packet weight and make easy to travel data.</td>
<td></td>
</tr>
<tr>
<td>Prediction based</td>
<td>+ Here, on the basis of frequently sensed data it is able to save time and energy by eliminating unusual data processing epocs.</td>
<td>- Needs the prior knowledge of explored domain for better modeling of expected value.</td>
</tr>
<tr>
<td></td>
<td>+ Some approaches have been proposed to run in light weight manner.</td>
<td>- Needs regular updation and validation of model to avoid deterioration in predicted values.</td>
</tr>
<tr>
<td></td>
<td>+ Great energy saving.</td>
<td>- Application specific.</td>
</tr>
<tr>
<td>Correlation based</td>
<td>+ Reduced number of nodes transmitting information, decreases contention in wireless medium.</td>
<td>- Prediction may abdicate sensed data accuracy.</td>
</tr>
<tr>
<td></td>
<td>+ These techniques are best in results when integrates with prediction and clustering.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Great energy saving.</td>
<td></td>
</tr>
</tbody>
</table>

In [26], authors proposed data gathering algorithm designed for network with high spatial and temporal correlation with the objective to reduce communication between source and sink. Result shows that the proposed algorithm is efficient when there is high spatial and temporal correlation. In [27], author presents the Clustered AGgregation (CAG) algorithm. In CAG, only one sensor reading per cluster is transmitted whereas with Tiny AGgregation (TAG) all the nodes in the network transmit the sensor readings.

<table>
<thead>
<tr>
<th>Application</th>
<th>- best suitable for the applications where occurrences of sensed data are very frequent e.g. Environmental monitoring.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging schemes</td>
<td>PINCO[19], Data Funneling[20], Entropy-based Compression algorithm[21]</td>
</tr>
<tr>
<td>Prediction based</td>
<td>- Prediction using variable step size LMS algorithm [23], Neural data prediction [24].</td>
</tr>
<tr>
<td>Correlation based</td>
<td>- Energy Efficient Spatial correlation [26], CAG[27].</td>
</tr>
</tbody>
</table>

In [26], authors proposed data gathering algorithm designed for network with high spatial and temporal correlation with the objective to reduce communication between source and sink. Result shows that the proposed algorithm is efficient when there is high spatial and temporal correlation. In [27], author presents the Clustered AGgregation (CAG) algorithm. In CAG, only one sensor reading per cluster is transmitted whereas with Tiny AGgregation (TAG) all the nodes in the network transmit the sensor readings.

<table>
<thead>
<tr>
<th>Application</th>
<th>- These are best suitable for event based applications e.g. Structural health monitoring applications etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belonging schemes</td>
<td>Energy Efficient Spatial correlation [26], CAG[27].</td>
</tr>
</tbody>
</table>

3.2.1 Sink mobility based approaches

These are mobile nodes works as destination of messages originated by sensors i.e. Sink is the endpoints of data collection in WSNs. Sink can either autonomously consume collected data for their own purposes or make them available to remote users by using a long range wireless Internet connection. In these cases, ordinary sensor nodes are static and densely deployed in the sensing area and single or multiple Mobile Sinks (MSs) move throughout the WSN to gather data coming from all ordinary sensor nodes. The path between the ordinary sensor nodes and the MSs is multihop. As the position of the MS is not fixed (i.e. sink nodes can be attached to people, vehicles or animals that can move inside the region of interest), the actual path changes with time. In [28], authors proposed dynamic tree reconfiguration protocol (DTR). This protocol maintains a temporally-sensitive tree which is updated via counts and most recent actions. In [29], authors proposed a network coding scheme called separate network coding (SNC) for data collection in WSNs with a mobile Base Station. This scheme provides efficient storage method for collected data within one time interval. This method introduces the data encoding, storage and decoding. Decoding is held at mobile base station. In [30], author proposed and evaluates a Mobile clustering algorithm (MCA) for periodical data gathering applications in WSNs. The algorithm partition the adjacent nodes which sense similar target into cluster using genetic algorithm, then through similar data aggregation, the amount of redundant data transmission is minimized.

3.2.2 Sensor node mobility based approaches

3.2.2.1 Mobile relays

Mobile relays (MR) are supportive nodes which gather messages from sensor nodes, store them, and carry the collected data to sinks or base stations. These are not the end points of communication rather they act as mobile forwarders. In short, the collected data move along with them until they reach in the range of sink or base station. In [31], authors proposed the data-MULE system which is a MR-based data collection system for WSNs. The data-MULE system consists of three-tier architecture; the middle tier is represented by relays known as mobile ubiquitous LAN extensions (MULEs). In [32], authors proposed a scheme which uses mobile gateway as mechanical data carrier which periodically travels the network to gather the sensor data; this tour starts and ends at the sink.
3.2.2 Mobile peers
Unlike MDCs, which are either sinks or special relay nodes, mobile peers are ordinary mobile sensor nodes in WSN-MEs. Since they can be both originator and relay of messages in the network, their interactions are symmetrical because the sink itself might be mobile. When a peer is in the communication range of the base station, it transfers its own data and data gathered from other peers during audit in sensing area. When WSNs comes in use of specific applications like environment monitoring (which includes forest fire detection, monitor disaster area or animal tracking etc.), military specific applications (enemy activity tracking, attack detection, battle field damage detection etc.) or many more commercial applications, it make sense if and only if sensor nodes are location aware. In [33], authors proposed a scheme for urban sensing in which people act as mobile peers to collect data from other portable devices or from the surrounding environment. In this case, sensor nodes are not used mainly for monitoring the environment, but to distinguish people in terms of both interactions and control information. Zebranet [34] and SWIM (Shared Wireless Infostation Model) [35] projects focus on wildlife monitoring applications by using mobile peers. Sensor nodes are attached to zebras in the Zebranet concept, or whales in SWIM system, so that they exchange the gathered information during encounters.

3.2.2.3 Mobile Relocatable node
These nodes are mobile nodes which change their location to forward data from the source nodes to the sink. These nodes do not carry data as they move in the networks; in fact, they only change the topology of the network. More specifically, after moving to the new location, usually they stay fixed and just forward data along multihop paths. To reduce communication holes, Scan-based Movement-Assisted sensoR deploymenT (SMART) algorithm based on Hungarian method has been proposed in [36]. This is also known as seed planting method. This method moves a sensor to each uncovered area to cover the holes. The results show that a cost effective sensor deployment can be achieved by using this SMART algorithm. In [37] authors proposed predefined, intelligent, lightweight iTopology managementT (PILOT): A system with relocatable nodes targeted for topology management. In this, special nodes are used to reestablish network connectivity if any fault occurs. In short, PILOT nodes act as bridges and they actively change the WSN topology in order to improve communication reliability.

3.2.3 Agent mobility based: Mobile agent
The mobile agent (MA) called a mobile data collector (MDC) tours the network and collects the data from the nodes. These MA basically self-contained software component that migrates from one host to another for the purpose of executing application specific functions. MA originally proposed to provide computational flexibility by carrying the processing logic to the data sources. These agents can also help in data processing, data aggregation and other functions if needed. In [38], author proposed Mobile Line Based Data Dissemination (MLBDD) protocol. This protocol uses a mobile agent to collect data by using aggregate functions to reduce the amount of data transmitted. The MA consists of a process code and a state; it resides on parent machine and dispatched to run on server (i.e. remote host), after completing its assigned task at the host, the MA migrate to another machine. In [39], authors proposed a clone-based dynamic and distributed agent migration (CDDAM). CDDAM consists of two phases: construction of rooted spanning tree phase and clone based agent migration phase. CDDAM performs the migration of agents within the network by creating virtual infrastructure of rooted spanning tree.

**TABLE III. Classification of Dynamic Data Processing Approaches**

<table>
<thead>
<tr>
<th>Approach</th>
<th>Features</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| Mobile sink based | + These are mobile nodes which are the destination of messages, originated by sensor nodes.  
+ Gains advantage over Hot Spot Problem.  
+ Optimized energy efficiency, lifetime and peer to peer delay achieved.  
+ Better to use with fixed network topology.  
+ More secure than static sink. | - Requires mobility control and mobility aware duty cycle management mechanism for better energy utilization.  
- Requires additional operations with routing like topological updates, contact detection etc.  
- Increased message delivery latency.  
- Increased packet loss ratio due to unavailability of fixed contacts. |
| Application: Deadline based application (animal tracking, battle field surveillance), forest monitoring applications etc. | Belonging schemes: DTR[28], MCA[30], SNC[29] |
| Mobile relay based | + These are supportive nodes which gather data from sensor nodes and carry the collected data to the sink.  
+ These are not the end point of data collection and not an originator of data; they simply act as a data relay.  
+ These mobile relay nodes helps in energy efficient data aggregation and dissemination in the wireless sensor network. | - Requires additional security approach for securing network from threats because intruders may introduce own node which may act as a relay; in such case security may suffer. |
| Applications: Military applications, monitoring applications etc.: Where sensors deployment is sparse; it may be very helpful for good connectivity and easy information exchange. | Belonging schemes: data-MULEs[31] |
| Mobile peer       | + These nodes can be both originator and relay of messages.  
+ Peer can be an animal, handheld device, vehicle, human etc. | - Long execution time and high overhead. |

*International Journal of Computer Applications (0975 – 8887) Volume 126 – No.11, September 2015*
4. CONCLUSIONS
We have outlined the static and mobile WSNs briefly and presented a detailed classification of their respective data processing approaches. The summarized information of each category is provided which can help researchers to select the most appropriate approach at the time of designing any scheme/protocol. If we talk about architecture driven data processing approaches in static WSNs, researcher’s most favorite approach is clustering. The outcome of the study is that the tree and cluster based approaches are best in terms of performance as they are easy to implements and most suitable in number of applications. The data driven approaches attracts researchers according to the need of application, i.e., whether it can compromise with data quality or not. The study concludes that mobile Agent approach is best suitable and highly preferable in dynamic data processing approaches because it includes intelligence in sensor networks which is software patch basically. Everyone is aware of the growth of software program in current time so when it is combined with WSNs becomes more advantageous. The comprehensive analysis made in this paper provides relevant information regarding suitability of particular approach for specific application.

5. REFERENCES


