Survey On Energy Efficient Data Collection In Densely Populated Wireless Sensor Network

M. Thangaraj¹, M. Renuga Devi², S. Anuradha³
¹,²,³(Department Of Computer Science, Madurai Kamaraj University, Madurai)

Abstract: WSN uses a large number of wireless sensor nodes to collect information from their sensing terrain. WSN is battery-powered devices in which energy-saving is a crucial issue. Every node in sensor network consist of three subsystems, first sensor subsystem which senses the environment, the second processing subsystem which do local computation on sensed data and third communication subsystem which is responsible for message exchange. In this paper, various literature was reviewed on different data collection mechanism on energy efficiency and presented the current scenario.

Keywords: Battery-powered, Data collection, Energy Efficiency, local computation, sensing terrain.

I. Introduction

Wireless sensor network is small, highly distributed, lightweight wireless sensor nodes [1]. Wireless sensor networks (WSN) consist of many sensor nodes inbuilt in it. Wireless sensor network has a range of applications including environmental monitoring, habitat monitoring, disaster management, security, military, etc. The wireless sensor network is densely deployed to give a wide coverage area. WSN uses a large number of wireless sensor nodes to collect information from the sensing terrain[2]. WSN is self-organized, low-cost and low power utilizing network. WSN is the nodes which can sense, compute and communicate the data[3]. The main characteristics of WSN are fault tolerance, scalability, production cost, hardware constraint, sensor network topology, transmission media and power consumption[3].

The main challenge of WSN is energy consumption. There is a lot of energy consumption while transmitting the data, but sensor nodes have limited energy. Wireless sensor network is an interconnection of nodes which are deployed for monitoring[4]. The factors needed to design the network should be considered such as coverage area, mobility, power consumption, communication, capabilities. A sensor network is composed of many of low cost sensor nodes, which are distributed in the system or near to it. The nodes which are very small and it consists of sensing, data processing and communicating components. The sensor nodes provide self organizing abilities inaccessible areas[4]. The challenges of energy awareness are proposed at all layers of networking protocols stack. The realization of sensor network needs the application of wireless ad-hoc networking techniques. The different techniques of data collection schemes and energy conservation techniques are seen in the Fig[1],[2].

Fig 1. Data collection Techniques
II. Data Collection: An Overview

Data collection is an important issue in wireless sensor network. Application of continuous data collection must use sensor nodes to continuously record environmental measures, such as temperature and target tracking and thereby send them back to the sink node [5]. Data coming from multiple sensor nodes are collected to reproduce the evolving process of certain phenomenon in the monitored area.

2.1. Complexity of Data Collection

In wireless Sensor Network often the goal is to collect the raw data or in-network processed facts to do some further analysis of sink nodes. The common many-to-one communication pattern used for the sensor network application is converge cast, which includes the basic complexity problems i.e., data collection, distributed data choice[4].

The complexity in data collection is divided based on its message, energy and time[4]. In order to decrease the energy consumption, the smallest energy cost path must be taken from its origin.

2.2. Scheduling in data collection

In order to make the freshness of data, the data should be collected quickly as possible. To do this the network should be able to manage the difficulties like delay, overhead, re-transmissions etc.

Key aspects of scheduling are:
- Minimizing Energy Consumption
- Minimizing Scheduling Length
- Minimizing Latency Maximizing Fairness.

2.3. Overview of Wireless Sensor Network

WSN is a new type of emerging network, where it consists of many special features when compared to traditional networks such as internet, wireless mesh network [5]. The first step to be carried is, a sensor node after deploying it is expected to work for days or weeks without intervention. Since the nodes are powered by the attached battery, high efficient energy use is necessary in which it differs from internet and wireless mesh network[5]. The second thing to be noticed in WSN is, it is not expected to work through a long time. The working of the sensor node is seen in Fig[4]

- It senses the defined environment.
- It Processes the sensed data.
- Transmit the collected data.
- The last state is the idle state in which it waits for next sensing-processing-transmitting cycle[5].

![Fig 2. Energy conservation schemes](image)

![Fig 3. Sensor node working](image)
The two main features supported by WSN are fault tolerance, duplicate sensing. In order to support fault tolerance, a place is often covered by several sensor nodes. Similarly to avoid the duplicate sensing, the other nodes are kept in idle state while one node is performing the sensing operation. Further the energy consumption can be reduced by making the idle nodes to a dormant state, where some of the components like wireless radio, processing unit are turned off. The nodes are again in active state by using the internal timer when the next cycle comes [5]

III. Various Techniques Of Data Collection In Wsn

3.1. Plain Data Collection
The Plain data collection is a method in which each sensor node sends its measurement to a base node, by which data is stored and processed [6]. Nodes are deployed in the sensor network in the form of closely located area to increase the accuracy and reliability of the collected information, since sensor reading has to be propagated back to the base node over multi-hops were used the wireless communication channel [6].

3.2. In-Network aggregation
It reduces the communication between the nodes. The process is carried out according to the specific aggregation function by combining data from different sources [6]. only data that is enough is transmitted by using the traditional operations such as MIN, MAX, AVG, COUNT [6]. In this work the energy efficiency is achieved, to keep the network operational as long as possible. The Number of Transmissions required for data collection is also done. [7]
It involves the full (centralized) Knowledge of the data correlations structures in which it is used to express the constraints for all the nodes. The second drawback is that it involves the calculation of SPT. [7] These are the issues which are not resolved.

3.3. Queries-Based Data Collection
The Query-Based data Collection is comprised of three phases,
- Query propagation phase
- Data Aggregation Phase
- The subsequent Response propagation phase

QBDCS mainly focuses on the routing mechanisms for sensor nodes, which forward the query and response packets. To minimize the energy consumption, the estimation of the pack delivery velocity and to predict the position and time when it meets the response packet [8]. In this work QBDCS is very effective, since it chooses the optimal query opportunity to send the query packet by tailoring the routing mechanism for partial node participation in WSN. The performance by minimizing the energy consumption and delivery latency in QBDCS is achieved.

Certain speed querying is needed to query in a specific area or in a point of interest to collect the sensed data. This is the major problem seen, where a position prediction technique should be used in advance. These are the issues which are not resolved.

3.4. Multipath data collection
The multipath data collection is used to overcome the high loss rates. The concept is if we limit the number of messages that each node transmits and receiver, the required level of energy efficiency can be achieved. The data propagation path from the sensor nodes is reduced in this protocol [8].

3.5. Feedback-Based Data collection
The process of data collection in WSN is based on feedback mechanism is from the source node to the sink. If the packet arrives at the sink at the end, which means that there is no node along the path. In order to improve the quality of the data collection is the historical information is stored [8]. In this work there is no loss of packets seen as because each time when the packet is arrived, feedback is generated.
- Large overhead.
- Weak wireless links.
- Insecure, [9]

The above mentioned points are the issues which are not resolved.

3.6. Optimal and suboptimal aggregation
The suboptimal schemes consist of three sub phases which are used to collect data in WSN. They are described in the following subsections.
3.6.1. Center at Nearest sources (CNS)
The most simplest phase in which the source sends their data directly to the source in which the source sends their data directly to the source in which it assumes the sink which is nearest to it and it sends an aggregated information to the sink[8].

3.6.2. Shortest paths Tree(SPT)
In this data collection method sends its information to the sink by the selected shortest path and if there is a overlapping path, it combines it to form the aggregation tree.

3.6.3. Greedy Incremental Tree (GIT)
It is the sequential type of scheme in which at the first step the aggregation tree with shortest path mechanism is done. After that step the next source is selected which is closest to the current tree is connected to the tree[8].

3.6.4. No Data Aggregation
In no data aggregation process, sensor devices should be unaware of the other neighboring nodes. Even though the data collected is small the sensor sense it and sends the data to the end nodes (sink). No data aggregation do not apply any data aggregation technique and it just simply forwards the data packets towards the sink node[9].

IV. Energy Conservation Schemes
The Three Main Components Available in data collection protocol are as follows

- How data collected in which a suitable collection function.
- The defined routing scheme in which it is characterized by the aggregated data routed towards the base station.
- The schedule of the data collection, which classifies the waiting time of the node before sending its data.

The energy conservation schemes in data collection are as follows

- Topology Control.
- Tree Based.
- Cluster Based.
- Sleep/Wake up Protocols.
- A Contention based protocols

4.1. Topology control
Network redundancy is the main characteristic seen in topology control Protocol. The mechanism is that, it can easily put to a low power consuming state in the case that when the sensor nodes are not in direct use[11]. The topology control is divided into connectivity driven protocol and location driven protocol activation and deactivation of sensor nodes is based on connection driven protocol[11]. Based on the location, whether the node should be turned on or not, is categorized Under location driven protocol[11].

4.2. Tree based topology
In the tree base topology the well known LEACH protocol is described. Efficient data collection and energy savings are achieved through these construction mechanisms[11]. Since it is not done in random manner it focuses the remaining energy.

4.3. Cluster based topology
The Network is divided into a number of clusters. Each cluster in this type consists of a cluster-head, which is selected among cluster members[11]. The cluster member sends the collected data and delivers the aggregated Data to the sink. The major factors when applying a cluster based data aggregation in a network. They are as follows.

- How many clusters to be formed to optimize the network performance parameters.
- How many numbers of nodes should be taken to make a single cluster.
- Which node is to select as the cluster head

4.3.1. Enhanced Low Energy Adaptive Clustering (ELEACH)
In this type of approach it consists of a non-uniform starting energy level between the sensors under certain assumptions, the energy consumption required by the cluster-heads are obtained by the square root of the total number of sensor nodes.
4.3.2. Leach – Centralized (Leachc)

This type of approach utilizes a centralized clustering algorithm and study-state protocol. This data about the current location and its energy level transfers to its base station (BS) during the setup phase of LEACH-C. Using global data of the network in order to provide enhanced cluster that require less energy for data transmission[11]. In this work the LEACH protocol is able to enhance the network lifetime.

- Data is sent to the base station at every round so energy consumption is high[13]
- The cluster heads are elected randomly, so the optimal number and distribution of cluster heads cannot be ensured[14]

The above mentioned points are the issues which are not resolved.

4.3.3. Hierarchical PEGASIS

Similar to PEGASIS, called hierarchical PEGASIS is used to decrease the delay occurred while transmitting the packet to the base station. In order to avoid collisions, simultaneous transmissions of data are studied that incorporate signal coding and spatial transmission. Simultaneous transmissions of data messages are used in order to reduce the delay which occurred in PEGASIS[11].

The main objective is to decrease the delay which incurred for packets during transmission to base station. The solution to the data gathering problem is to introducing the energy x delay metric.[15][16].

4.3.4. Energy Balancing PEGASIS(EB-PEGASIS)

EBPEGASIS is a type of energy based efficient chaining algorithm. The node itself will assume that the closest node is a far node[11]. So merging process is done by the closest node, which then it will emerge in a long chain. In this work EB-PEGASIS avoids this phenomena using a distance threshold. It not only saves energy on the threshold, but also balances the energy consumption of all sensor nodes[12].

4.4. Contention Based Protocols

In a wireless sensor network, contention based protocols are used which are the MAC protocols. B-MAC (Berkely MAC) is the most efficient protocol, which is a contention based which has low complexity. Channel access and most important functionality like energy efficient process are provided by B-MAC[11].

Most common MAC protocols are
- TDMA (Time Division Multiple Access) based
- Hybrid Based
- Contention-Based

4.5. Other type of data collection

These data collection techniques are analyzed in terms of cost, reliability and energy.

4.5.1. Distributed Data Aggregation Protocol (DDAP)

It is a self-organizing data collection protocol, which distributes the data aggregator roles among the sensor node randomly. Since there is no central authority, it is simple. In upcoming surveys, more powerful adaptive DDAP can be proposed for efficient power saving[18].

4.5.2. Power-Graded Data Gathering [PODA]

The energy efficient and new data gathering mechanism are power gathering (PODA) [18]. In order to gather data in large areas and to tackle the hotspot problem, the output power is adjusted at the system level. Nodes in the network consume the energy in an even manner. The lifetime is prolonged and energy is improved. In this work the whole network the consumption of energy is even. The protocol is simple as because it is having a low protocol cost and easy implementation.

- The Energy balancing problem is not considered when the sink node is not at the center of the network.
  
  The above mentioned issue which are not resolved

V. Different Energy-Efficient Techniques In Data Collection

5.1. Clustering based lifetime maximizing aggregation tree

CLMAT is a technique in data aggregation with the aim to reduce energy consumption, minimizing the cost in terms of energy consumption. The node which has a maximum available energy in CLMAT is used as parent node aggregator node[19]. By using these parameters the network lifetime can be enhanced.
5.2. Polynomial regression based secure data aggregation

PRDA is a protocol in which the sensor nodes represent their sensed data in terms of polynomial functions. In order to preserve the privacy of the data collected, PRDA protocol can be used by employing the polynomial regression on sensor data series.

5.3. Grid-based Architecture

The Grid-based architecture is an energy efficient data storage scheme in which the network is equally divided into two dimensional logical grids in which the grid is denoted as N. The process carried out in this is based on active and sleep mode procedure. ie., If one sensor is in sleep mode at one time slot and the other one is active[19].

Since the grid based architecture is of distribute one the merits in grid-based architecture are,

- Computation complexity
- Transmission delay required for the computations
- Deployment / Reconfiguration[20].

5.4. Temporal correlation based data aggregation scheme

The ARIMA model which is also called Box-jenkins model is a widely used forecasting model for data aggregation. It is based on periods updating of time. If it is high, the sensed value is saved in the sensed value to the aggregator. [19]. Energy efficiency can be achieved by done by checking the periodically collected data using the ARIMA model.

5.5. Adaptive approximate data collection for wireless sensor network

ADC is a type of data collection in which the sensor network is divided into clusters and also by discovering local data correlations on each cluster head [21]. In this work, by reducing the data traffic within a sensor network, the packet delivery ratio can be greatly increased.

5.6. Fair data collection scheme in wireless sensor network

The slow congestion detection and rate convergence problems can be solved using fair data collection[22]. The Data control can be reduced by improving the rate of the congested node sending data. In this work the congestion will reduce the throughput of WSN.

5.7. Energy-Efficient Wake-up scheduling for Data collection and aggregation

Energy-Efficiency and free channel access can be reduced by TDMA-based Wake-up scheduling algorithms are also used for message delays by using consecutive time slots at different radio states, the number of state transitions can be reduced with sensor nodes. In this work the energy consumption is reduced.

5.8. A hierarchical scheme for energy efficient data collection

5.8.1. The hierarchical clustering scheme

The Whole WSN is divided into a set of smaller network clusters in the hierarchical approach. The Base station is chosen from the center node[23]. Here the data collection of the base station is performed in two phases. The data are collected from all cluster heads, which is in their own clusters. The data are aggregated in their own clusters. In the second phase, the WSN’s base station collect data from all the cluster heads[24].

5.9. Low Energy Adaptive Clustering Hierarchy Protocol

In the LEACH (Low Energy Adaptive Clustering Hierarchy Protocol), the sensor nodes are grouped into small clusters. One of the energy potential members of the group have been elected as the Cluster Head (CH), the other nodes act as a Cluster Member (CM). In FTEP (Fault Tolerant cluster head Election Protocol), there is a possibility of rotating the cluster head based on the energy level after the potential energy scans by the distributed-dynamic cluster head election algorithm. In a multi-hop manner the cluster head collects the data from the cluster members. By this process of organizing the nodes into clusters, the long distance of communication is avoided, thus we save the energy[25]. There are cases where the local data are fused by the data compression techniques for minimizing the bits of data transmission from the cluster head to the gateway node. In EEMC (Energy Efficient Multi Level Clustering), the cluster head selection is based on the probability factor, minimum latency in communication[26]. Here the data is collected in multiple rounds on the remotely distributed sensors to the cluster and cluster to the gateway. With the PACBR protocol (Power Aware Cluster Based Routing), the communication between two nodes are governed by the inverse square law of energy.
5.10. Data Dissemination in the grid based routing

To achieve energy efficiency, the virtual grid is constructed based on the Steiner Points among the nodes and on event detection, through the grid structure the data is disseminated. In TTDD (Two Tier Data Dissemination), the nodes in the cross section of the grids plays the role of the sink and responsible for data transmission. With the EEDD (Energy Efficient Data Dissemination) routing, the target location aware and target area aware schemes, routes the packets in a more energy efficient manner. For VGDD (Virtual Grid Based Data Dissemination), the virtual grids are formed along with the communication rules for data propagation[25]. The cell header readjusts and determines the route selection. CHESS protocol considers the node’s energy cost of forwarding to choose the next hop for forwarding.

5.11. Tree Construction and Data Forwarding

Collection Tree Protocol (CTP), the tree topology is formed among the nodes in the WSN. Each node establishes the synchronous neighboring table with the synchronization points. The node transmits data with the SYNC preamble to inform the synchronization details for the data propagation[20]. Multi channel CTP, helps the node to overcome the low interference by defining the new routes. CTP considers the agile link estimation, data path validation, an adaptive control beacon scheme to define the path. Data Mules are defined first and it is assigned to the travelling salesman problem to find the shortest path to reach the sink or gateway. It is energy efficient in smaller networks. MST (Minimum Spanning Tree), SST(Shortest Spanning Tree) are some of the existing tree structures for the energy efficient data collection topologies considered.

5.12. Time Sensor Data Collection Using Distributed Graph Coloring – TIGRA

With the TIGRA transmission scheduling algorithm, there are three distinct phases. First one is round determination where the node wants the data from its children – the children determine which epoch it has to participate[27]. The second phase is the slot determination phase where the slot within the round is decided for different nodes for transmission to avoid interference and maximum utilization of channels. In the third phase of data collection, on the determined round/slot/epoch each node transmits and the parent node collects the data.[28]

VI. Conclusion

In this paper, the investigation of WSN basics, applications of WSN such as military, environmental, health and other commercial applications are seen. The various techniques of data collection in WSN and design of data aggregation are formulated and then the working of WSN is seen with two different phases. The energy efficiency is one of the main challenges in the design of protocols for WSNs. Due to the scarce energy resource of sensors. The vital objective behind the data collection algorithm is to operate sensors as long as possible, to extend the network lifetime. The survey work mainly focuses on the energy efficient data collection algorithms for WSNs.

References

[3]. Feng Wang, Student Member, IEEE, and Jiangchuan Liu, Senior Member Networked Wireless Sensor Data Collection: Issues, Challenges, and Approaches, submitted to ieee communications surveys & tutorials for possible publication, pp 1-2.
Survey On Energy Efficient Data Collection In Densely Populated Wireless Sensor Network

[14]. k. dascrewman, The University of Texas at Arlington and giuseppe anastas, Data Collection in Wireless Sensor Networks with Mobile Elements. pp 73-79


[22]. Mousam Dagar1 and Shilpa Mahajan21,Data Aggregation in Wireless Sensor Network: A Survey , 2Department of CSE/IT, ITM University, Gurgaon, India,pp 10-17.


DOI: 10.9790/0661-1802044855 www.iosrjournals.org 55 | Page