ISBN 978-93-5156-328-0
International Conference of Advance Research and Innovation (ICARI-2015)

Data Collection Method to Improve Energy Efficiency in Wireless Sensor Network

Khushboo Gupta [a], K. P. Yadav [b]
[a] Department of Computer Science & Engineering, Uttar Pradesh Technical University, Lucknow, U. P., India
[b] Department of Computer Science & Engineering, Mangalmay Institute of Engineeering and Technology, Greater Noida, Uttar Pradesh, India

Article Info
Article history:
Received 3 January 2015
Received in revised form
10 January 2015
Accepted 20 January 2015
Available online 31 January 2015

Keywords
Wireless Sensor Networks,
Energy Efficiency,
Data Collection,
Clustering,
Aggregation Tree
Network Coding,
Correlation Dominating Setetc

Abstract
Wireless Sensor Networks (WSNs) are generally self-organized wireless ad hoc networks which incorporate a huge number of sensor nodes which are resource constraint. Among the tasks of WSN, one mostessentialtask is to collectthe data and transmits the gathered data to a distant base station (BS). The effectiveness of WSNs can be calculated in terms of network lifetime.Data collection is a frequentoperation but analytical and critical operation in many WSN's application. To prolong network lifetime innovative technique that can improve energy efficiency are highly required. This paper presents a survey for designing Energy Efficient Data Collection Methods used for prolonging network lifetime in Wireless Sensor Network (WSN). The study highlights the importance of different Data conditions for various purposes like emergency response, medical monitoring, military applications, surveillance in volcanic or remote regions, etc.Different Data Collection methods like data aggregation clusters, data aggregation trees, network coding, correlation dominating set, etc. are considered in detail in this study. Furthermore a comparison of different Data Collection Method based on the network lifetime, energy efficiency, complexity of the algorithm, transmission cost and fusion cost is done.

1. Introduction

A WSN is a set of distributed independent sensors which are capable of monitoring the environmental and physical conditions and it collectively transmit the data to a BS. The development of WSN was motivated by the military applications and presently WSN are used in weather monitoring and forecasting, forest fire monitoring, industrial applications, health monitoring, agriculture, physical environment monitoring, and so on. [2] - [4]

In addition to the unreliable wireless communication, WSN have to work with limited resources such as limited energy, limited processing capacity, limited storage, limited memory, limited communication capacity, etc. So, WSN suffers from lot of problems such as lost messages, delay in data delivery, data redundancy, loss of data, etc. [4] - [6], which results in low quality of service (QoS). Energy availability is the most important challenge in WSN. The data collected by the sensor node have to reach the sensor gateway at the earliest. Different energy efficient methods in WSN have to ensure anon going and continuous network and reduced power consumption. Changing the limited power sensors is inefficient in almost all applications. Due to which, the battery lifetime determines the network lifetime. So in WSN energy conservation and energy controlling have so much importance that it manage and control the energy efficient techniques and protocols.

The data collection methods can be categories into various types such as aggregation trees, cluster formation, network coding, correlation dominating set, etc. [1]. To increase network lifetime and scalability, clustering can be done which is an effective topology control approach. In cluster formation technique, cluster heads are responsible to transmit data to the BS and the remaining nodes are attached to the cluster head.

Corresponding Author,
E-mail address: khushboogupta2806@gmail.com

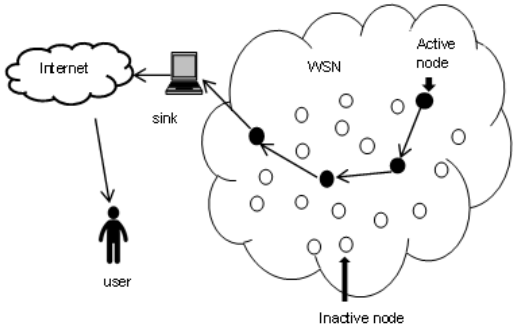

Fig: 1. Active Nodes in a Wireless Sensor Network

Aggregation is the process of merging data from different sources together. The aim of Data aggregation protocols is eliminating redundant data transmission and thus improving the network lifetime WSN. To combine the data in network aggregation trees are formed and the correlated data which is collected by the sensor nodes are combined and transmitted. Many aggregation methods are there and two methods are discussed in this paper.

In network coding technique is the global process of collection and routing information by a multi-hop network. With the goal of minimizing energy consumption, this technique processes the data at the intermediate node and thereby increasing network lifetime. There are different coding techniques which will decrease the transmission cost.

In correlation dominating set technique few sensor nodes are activated so that they can sense the data. By using

the correlation of data which is gathered by the nodes, the active nodes will also collect data for the remaining inactive node. The active nodes will transmit the data to the central node by forming a tree.

The following section consists of the explanation of these techniques and comparisons based on the advantages and disadvantages of these techniques are presented.

2. Data Collection Techniques Based On Clustering

2.1 LEACH (Low-Energy Adaptive Clustering Hierarchy)

The LEACH is a self-configuring and an adaptive cluster formation technique. LEACH is a clustering based protocol that uses data fusion to reduce the amount of data transmission. LEACH protocol was a major improvement over conventional clustering approaches in WSN. LEACH protocol is completely distributed in the sense that it does not need control information from the BS and nodes do not require global topology information.

In this technique, the nodes arrange themselves into local clusters and select one node as a cluster head. This selection of the cluster head is done on the basis of the total system energy and the nodes residual energy. The non-cluster head nodes can sent request to the selected cluster head to join. The cluster head collects data from the non-cluster head nodes to perform signal processing functions and transmits the collected data to a BS. The balancing of node energy usage is done randomly by rotation of cluster heads. Thus LEACH protocol has two phases: Setup phase and Steady phase. In the first phase the selection of cluster-head is done and non-cluster nodes are added to the cluster head. In the second phase the cluster heads collects and transmits the data to the base station [1].

The disadvantage of this technique is that it limits the scalability of the nodes in network. The non-cluster nodes have to send data in the allotted time. The cluster head may not be uniformly distributed over the network. The exact correlation is presumed which may not be corrected in some cases.

2.2 HEED (Hybrid and Energy-Efficient Distributed Clustering)

For long lasting WSN, the HEED protocol lis a distributed clustering protocol. HEED protocol does not made assumptions about the infrastructure or node processing abilities, it emphasis on different energy levels in sensor nodes.

HEED [7] protocol periodically selects cluster headson the basis of two parameters from the sensor nodes. These parameters are the node residual energy and the node degree or node closeness to its neighbours. HEED protocol ensures the connectivity of the clustered network as it makes required boundaries on intra-cluster and inter-cluster communication ranges and on node density. This approach is hybrid: cluster heads are selected probabilistically based on the above defined two parameters and nodes join these clusters so that transmission cost is reduced. This approach supports a scalable data aggregation and effective in increasing network lifetime. It spends, spends less energy in clustering and can provide features such as load-balancing, less overhead. It is a complex algorithm as it has multiple iterations. Due to large number of iterations there are only little chances to decreases residual energy Advantages of this method are Scalable data aggregation, Prolong network lifetime and Load balancing. The disadvantage of this method is the complexity of the algorithm.

2.3 CAG (Clustered Aggregation Technique)

CAG technique forms node clusters by sensing similar values within a defined threshold based on the spatial correlation. The clusters remain same or unchanged until the sensor reading stay within a threshold over time based on temporal correlation. Since in CAG technique, every cluster transmits only a single reading, it is efficient in terms of energy. The result provided by CAG is an approximation aggregation with negligible and bounded error.

By using the correlation of sensed data, CAG minimizes the number of transmissions and aggregate queries to provide approximate results. This result is assured to be within an error-tolerance threshold provided by the user. Whenever a query is circulated in the network, the clusters are formed based on the correlation. Thus CAG technique has 2 phases: Query phase and Response Phase. In Query phase the data is transmitted by a tree formed. In the Response phase, the clusters head transmits data to the next level of the tree after aggregation [9].

CAG has various advantages such as it is energy-efficient in-network aggregation. It has leveraging both spatial and temporal correlations and is resilient to the packet loss .It ensures a confined approximation. CAG has certain disadvantages as thus technique uses simple correlation in which the selection of the cluster-head is done by the edges of the forwarding tree and connecting sensor nodes and no hybrid clustering protocol.

3. Data Collection Techniques Based On Data Aggregation Trees

3.1 Scale Free Aggregation

In this technique, an efficient tree is constructed for transferring data to a single central node for processing by the different sensor. Data from different node passes to a central node and then data can be easily aggregate in that node. This technique reduces the energy consumption and transmission cost and will increase the network lifetime.

There are two types of aggregation methods: single sink aggregation and multiple sink aggregation. In first method data from different sources will aggregate through a single link by a aggregation function [13][14].This method lay emphasis on development of a data aggregation tree that can work for all types of aggregation functions. Single sink aggregation has two steps, matching step and selection step. By using these processes the single link can be found and a aggregation tree is formed. A single sink method, a single tree is used with all the aggregation functions. But this it is not useful when the aggregation function depends on the source identity and also when multiple sources and sinks are present in the tree.

3.2 MFST (Minimum Fusion Steiner Tree)

This is a routing algorithm for energy efficient data collection and is based on the techniques presented in [10, 11]. This method works on the defined metrics to pairs up the source nodes and then randomly chooses a center node

from the node-pairs. By paying appropriate transmission and fusion cost on an edge, the weight of the non-center node will be transferred to the center node. Afterwards, on-center node will be eliminated and the center node will be aggregated as a new set of sources with the combined weights. This process is continued on the new set until the sink nodes is the only remaining node.

For collecting the data MFS uses fusion which incorporates optimizes over the data transmission cost and the cost for data fusion. Thus in MFS the data fusion cost as well as the transmission cost is also minimized, which is essential for critical data and for security of WSN. The fusion cost is determined by the fusion function by using the input of the function. If a complex fusion function is used then the fusion cost should be decreased.

MFS technique uses step by step process for data aggregation in that a single input is aggregated to the input set which already exists. Fusion cost plays a vital role for making routing decisions at the time of data aggregation. Due to the limited resources availability and variation in inputs arrival time, the transmission cost depends on both the amount of data being transmitted and on the cost of transmitting data by the link. The fusion cost depends on the algorithm used for fusion and amount of data to be fused [12]. Both fusion and transmission costs are calculated by this technique get an energy efficient method for data.

4. Data Collection Techniques Based On Network

4.1 On Network Correlated Data Collection

In on network data collection uses two coding strategies for minimizing the transmission cost of data from source to BS. There two coding strategies are Slepian-Wolf model and joint entropy coding model. These models differ from one another in optimization of transmission and complexity of optimal coding. These methods are used when there is a correlation between the data sensed by different nodes. In Slepian-Wolf model the transmission optimization is easy but coding optimal coding is very complex. The nodes near the sink have high data load whereas the nodes away from the sink have mower data rate. In case of joint entropy coding the optimal coding is simple but the transmission optimization complexity is high. Here the nodes far from the have data rate whereas nodes closer to the sink have lower data rate [16]. A combination of both these methods is more beneficial and effective.

4.2 Single Input Coding

The single Input Coding increases energy efficiency by correlating between the data. With the objective of minimizing resource consumption the data can be processed at the intermediate nodes. This is a global process of data collection and increase network lifetime. Single input coding is applied to the asynchronous network with no time consideration [15].

This technique determines the interrelationship in data sensed by various sensors to combine data from different sources. Two types of coding can be used: foreign coding uses MEGA and self-coding which uses LEGA. Conditional coding, In-network data processing, self-coding and foreign coding are also used to aggregate data. Conditional coding is one in which data from one source is combined in the presence of data from a other sources whereas In-network data processing is used when data goes from source to sink. In foreign coding data is coded in a foreign node whereas inself-coding data is coded in the originating node.

5. Data Collection Scheme Based on Correlation Dominating Set

5.1 IAND (Iterative Active Sensor Node Determination)

In this technique a small subset of sensor nodes is selected which should be enough to reconstruct data for the whole sensor network. During the gathering of data these selected active sensor nodes have to involve in the communication [1]. Since there is a need to relay data to the data gathering node, the selected nodes should be connected. In order to determine the nodes which will take part in the transmission of data to the data collecting node are the residual energy of the sensor node and the correlation between the data sensed by the nodes.

6. Comparison of Different Data Collection Techniques

This section lists and discusses the merits and de-merits of different data collection methods in the table.The comparison of these techniques is shown in table I. The comparison is done based on energy efficiency, network lifetime, complexity of the algorithm, transmission cost and fusion cost.

Table: 1. Comparison of Different Data Collection Techniques

Methods	Advantages	Disadvantages
LEACH (Low-Energy Adaptive Clustering Hierarchy)	It is Completely distributed, Requiring no control information from the base station, Nodes do not need global topology information.	Nodes must have data to send in the allotted time, Assumed perfect correlation might not be true
HEED (Hybrid and Energy-Efficient Distributed Clustering)	It ensures Prolong network lifetime, Scalable data aggregation, Load balancing	It is a Complex algorithm
CAG (Clustered Aggregation Technique)	It provides Energy-efficient in-network aggregation, Leveraging both spatial and temporal correlations resilient to the packet loss, Ensure bounded approximation	No hybrid clustering protocol

Scale free aggregation	It ensures Prolonged network lifetime, Efficient data aggregation, Decreases amount of data transmitted	Multi-commodity version of the problem, Aggregation depends not just on the number of sources, but also on the identity of the sources
MFST (Minimum Fusion Steiner Tree)	It has Varying network topology, fusion cost, correlation	It is Less robust.
On network correlated data collection	It reduces Total transmission cost	It has a Complex structure
Single input coding	It ensures Efficient collection of data	It assumes Star topology, Only single input coding
Active node determination	It has High performance, Increased network lifetime, High runtime efficiency	It can be used only for Static sensor network

7. Summary

This paper is an investigation and analysis of different data collection techniques that can be used to increase the energy efficiency in a WSN. To improve energy efficiency different data collecting Techniques can be used and these techniques reduces the energy utilization and increases the network lifetime. The different data collecting techniques to improve energy efficiency are various data collection Techniques based on clustering, network coding, aggregation trees and correlation dominating set. The energy consumption can be reduced and network lifetime can be increased to agreat extent using these methods.

References

[1] E. Karasabun, I. Korpeoglu, C. Aykanat, Active node determination for correlated data gathering in wireless sensor networks, Computer Networks, 2013, 1124–1138

[2] F. Zhao, L. J. Guibas, Wireless sensor networks: Aninformation processing approach, Morgan Kauffman, San Fransisco, 2002

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Computer Networks, 38(4), 2012

[4] C. S. R Murthy, B. S. Manoj, Ad hoc wireless networks: architecture and protocols, Prentice Hall, New Jersey, 2004

[5] K. Akkya, M. A. Younis, Survey of routing protocols in wireless sensor networks, Elsevier Ad Hoc Network Journal, 3(3), 2005

[6] A. Boukerche, F. H. Silva, R B. Araujo, RW. Pazzi, Low latency and energy event ordering algorithm for wireless actor and sensor networks, in Proc. 8th ACM ISMASWMS, Montreal, 2005

[7] O. Younis, S. Fahmy, Heed: a hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks, IEEE Transactions on Mobile Computing, 2004, 366–379

[8] Y. SunHee, C. Shahabi, Exploiting spatial correlation towards an energy efficient clustered aggregation technique (cag), IEEE International Conference on Communications (ICC), 2005, 3307– 3313

[9] A. Goel, D. Estrin, Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk, 2010

[10] In Proceedings of ACM-SIAM Symposium on Discrete Algorithms, Baltimore, MD, 2003

[11] A. Meyerson, K. Munagala, S. Plotkin, Cost-distance: Two metric network design, in Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, 2000

[12] H. Luo, Y. Liu, S. Das, Routing correlated data with fusion cost in wireless sensor networks, IEEE Transactions on Mobile Computing, 2006

[13] R. Cristescu, B. Beferull-Lozano, M. Vetterli, On network correlated data gathering, IEEE Proc. INFOCOM, Hong Kong., 2004

[14] A. Goel, D. Estrin, Simultaneous optimization for concave costs: single sink aggregation or single source buy-at-bulk, in: SODA'03 Conference, 2003

[15] P. V. Rickenbach, R. Wattenhofer, Gathering correlated data in sensor networks, 2004

[16] R. C. Baltasar, R. Cristescu, B. Beferull-lozano, M. Vetterli, On network correlated data gathering, in: IEEE Infocom Conference, 2004