IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661 Volume 3, Issue 3 (July-Aug. 2012), PP 08-11
www.iosrjournals.org

Energy Efficient Data Mining in Multi-Feature Sensor Networks Using Improved Leach Communication Protocol

Shivanna K[1], Shalini Kumari H A[2]
[1,2]*Department of Computer Science and Engineering, Shridevi Institute of Engineering and Technology, Tumkur,*

Abstract: *Wireless Sensor Networks (WSNs) with large number of sensor nodes are used to monitor the sensing filed. Sensor node energy efficiency and sensor network data-transfer reliability are the primary design parameters. Some applications may require heterogeneous sensor nodes with different sensed phenomena and different hardware characteristics. This heterogeneity imposes an added constraint to the mining of useful information from the network. At the same time, a stream of data that is frequently reported from each node to the base station or sink may be needed. In this paper we are considering multi-feature sensor network where the sensors are having ability to measure more than one phenomenon. Our focus is on developing a new clustering algorithm that handles the concept of multi-feature sensor networks.*

Sensor nodes in WSN are based on energy, minimizing energy consumption is an important issue in WSN. Data mining in such multi-feature sensor network is also another important issue. As the existing LEACH protocol, the protocol used for clustering, is having drawbacks with respect to energy consumption. We are building new improved or modified LEACH communication protocol that consumes less energy compared with the existing one.

Index terms: *Wireless Sensor Networks (WSNs), Data mining, LEACH, HEED, TDMA, Multi-feature, Hierarchical Clustering.*

I. Introduction

A Wireless Sensor Network (WSN) consists of a large number of sensor nodes that co-operatively monitor a specific region of interest. Typically, a sensor node is a small hardware device consisting of a processing unit, a sensing unit, a communication unit and a power unit that is used for sensing, data processing and communication purposes. These nodes collectively gather sensed information and forward it to the special node called *base station* which acts as interface between the sensor nodes and users. The uniqueness of a sensor node lies in its small size and light weight. However, there are a lot of constraints such as limits on resources in terms of energy, memory, computational speed, band width and so on [1][2].

Because of these constraints the interactions between sensors are limited to short distances and low data rates. These sensors are used in wide range of applications and real time applications such as nuclear power plants, habitat monitoring, military applications, security purpose etc [3].

Sensors gather useful information in a timely manner and send it to a centralized node named *sink*. The sink node is also known as base station and is responsible for further processing such as node query. Due to the large number of sensor nodes and the voluminous data that should be reported, data communication should be done in energy efficient manner. Centralized solutions for data collection are not recommended. Obvious drawbacks of this type of solutions include hindering the network since the base station becomes a bottleneck, the bandwidth allocated is not efficiently used and all sensor nodes consume a lot of scarce and valuable power to communicate with the base station. Hence, the optimum solution becomes a distributed data collection algorithm, where data mining techniques such as clustering are applied to the sensor nodes. When there are large number of sensors in the sensing field, sensors will be clustered to reduce the data redundancy. The cluster head will take care of this work.

Clustering of sensor nodes is considered as one of the very successful techniques of mining useful information from a distributed environment. It is a particularly useful technique especially for applications that require scalability to hundreds and thousands of nodes. Clustering also supports aggregation of data in order to summarize the overall transmitted data. However, the current literatures either focus on node or data clustering alone. Clustering of sensor nodes deals with two main operations: 1) identifying cluster heads, and 2) assigning nodes to respective cluster heads. These two operations should be done at a very energy-efficient level.

On the other hand, data clustering deals with collecting the similar data for aggregation purposes. The process of choosing the cluster head should take into consideration node design factors such as energy level of the sensor node and load balancing as well as their similarity in terms of the sensed data. A successful clustering algorithm is the one that produces an optimal amount of clusters, with each having a single cluster head responsible for inter and intra-cluster communication.

This paper *focuses* on the problem of minimizing the energy consumption of sensor nodes during cluster formation and aggregating sensed data of the sensor nodes of the cluster. Minimization of energy can be achieved by using improved or modified LEACH communication protocol. Where the previous work uses LAECH protocol to form the clusters and to aggregate the sensed data of the sensor nodes of the cluster, but it is having a drawback in consuming energy of the sensor.

This improved LEACH protocol is used in data mining in multi-feature sensor networks where it is having ability to sense multiple phenomenons.

The rest of the paper is organized as follows: Section 2 reviews related work. Section 3 gives working of the LEACH protocol. Section 4 gives proposed protocol or improved LEACH protocol. Conclusions and future research topics are drawn in Section 4.

II. Related Work

The work [1], [2], [14] gives the working of the sensor nodes. The work in [4] gives routing paths estimation in WSN. Handling sensors data becomes more irritating when heterogeneous nodes with more than a sensed feature is considered. Heterogeneous sensor networks may contain two or more different types of sensor nodes. These nodes differ in many hardware characteristics, such as communication ranges and battery sources, as well as sensing capabilities. Opposite to the traditional homogeneous networks, heterogeneity adds a new constraint to the data collection process especially when multiple features are reported by a sensor node. The problem of clustering data is a well studied problem in the literature. It has been used for very large databases [5]. Another main use of clustering protocols has been investigated for ad-hoc networks such as in [6] and [7].
Sensor networks clustering algorithms have been proposed by several researches as well where the idea of organizing a WSN into clusters is studied in [8] and [9]. For instance, in [9], the authors use LEACH (Low-Energy Adaptive Clustering Hierarchy) as a randomized technique to rotate between cluster heads in order to preserve energy and distribute the load evenly across the nodes in the network. A more adaptive approach is used in HEED (Hybrid Energy Efficient Distributed Clustering) [10], where the cluster head formation is dependent on the energy level of the sensor node. In case of HEED, the authors argue that the algorithm yield more distributed clusters and is efficient in terms of processing.

Little attention has been given to clustering of sensor nodes according to data readings. For instance, Taherkordi et al. in [11] propose multi-dimensional clustering algorithm, in which nodes are clustered according to their sensed attributes. On the other hand, the Distributed, Hierarchical Clustering and Summarization algorithm (DHCS) proposed in [12] provide a better solution for dense networks. The algorithm adopts several techniques, such as difference and hop count thresholds to model node and distance-based clustering.

A very simple node clustering algorithm is presented in [7] named Local Negotiated Clustering Algorithm (LNCA). The algorithm clusters sensor nodes primarily on data similarity. A common drawback in all of the current clustering algorithms is that they consider a sensor node with a single reporting feature. This might lead to inefficient clustering in terms of sensors consumed energy as well as the sensors reporting reliability. In addition, up to our knowledge, there is no standardized data mining framework that simplifies the data mining process in sensor networks.

The work in [13] proposes a global sensor network framework, sensor network model and the authors suggested different algorithms that can be used in each layer. But this framework is having a drawback in energy consumption of the sensor nodes and in cluster head selection while forming the clusters. In this paper to overcome this drawback we are proposing a Multi-Feature Improved Leach Based Clustering (MFILBC) method.

III. Leach Protocol

Low Energy Adaptive Clustering Hierarchy (LEACH) is the first hierarchical cluster-based routing protocol for wireless sensor network which partitions the nodes into clusters, in each cluster a dedicated node with extra privileges called *Cluster Head (CH)* is responsible for creating and manipulating a *TDMA* (Time Division Multiple Access) schedule and sending aggregated data from nodes to the BS where these data is needed using *CDMA* (Code Division Multiple Access**)**. Remaining nodes are cluster members.

This protocol is divided into two rounds; each round consists of *two* phases:

A. Set-up Phase

Each node decides independent of other nodes if it will become a CH or not. This decision takes into account when the node served as a CH for the last time (the node that hasn't been a CH for long time is more likely to elect itself than nodes that have been a CH recently).

In the following advertisement phase, the CHs inform their neighborhood with an advertisement packet that they become CHs. Non-CH nodes pick the advertisement packet with the strongest received signal strength.
In the next cluster setup phase, the member nodes inform the CH that they become a member to that cluster with "join packet" contains their IDs using CSMA. After the cluster-setup sub phase, the CH knows the number of

member nodes and their IDs. Based on all messages received within the cluster, the CH creates a TDMA schedule, pick a CSMA code randomly, and broadcast the TDMA table to cluster members. After that steady-state phase begins

.

B. Steady-state phase

Data transmission begins, Nodes send their data during their allocated TDMA slot to the CH. This transmission uses a minimal amount of energy (chosen based on the received strength of the CH advertisement). The radio of each non-CH node can be turned off until the nodes allocated TDMA slot, thus minimizing energy dissipation in these nodes.

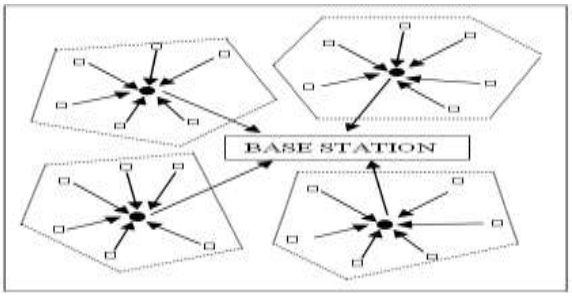

Figure 2: LEACH Protocol Architecture.

When all data has been received, the CH aggregates these data and sends it to the BS. LEACH is able to perform local aggregation of data in each cluster to reduce the amount of data that transmitted to the base station.

Although LEACH protocol acts in a good manner, it suffers from many drawbacks such like;

- CH selection is randomly, that does not take into account energy consumption.
- It can't cover a large area.
- CHs are not uniformly distributed; where CHs can be located at the edges of the cluster.

IV. Proposed Method And Protocol

There are two types of sensor networks which are single and multi-feature sensor networks. A single feature sensor network is a network with each sensor node reports only one feature. A multi-feature sensor network is a network with nodes report more than one feature. Further, the network could be classified into homogenous and heterogeneous sensor networks. In homogenous sensor networks, nodes are typical in every aspect. However, in heterogeneous sensor networks, sensors differ in their characteristics such as initial energy, sensing range, and communication ranges. In this paper, we are interested in multi-feature heterogeneous sensor networks. Therefore, we introduce MFILC as a new clustering algorithm that fits the purpose of multi-feature sensor networks. MFILC adapts improved LEACH clustering technique to support the multi feature sensor networks. To overcome the drawbacks of the LEACH method we are using improved LEACH clustering technique. We extend the LEACH protocol to multi-hop clustering in a multi-feature environment as well.

In our new version of LEACH protocol or Improved LEACH, the cluster contains, CH (responsible only for sending data that is received from the cluster members to the BS), sub-CH (the node that will become a CH of the cluster in case of CH dies), cluster nodes (gathering data from environment and send it to the CH), as shown in figure 3.

In the original leach, the CH is always on receiving data from cluster members, aggregate these data and then send it to the BS that might be located far away from it. The CH will die earlier than the other nodes in the cluster because of its operation of receiving, sending and overhearing. When the CH die, the cluster will become useless because the data gathered by cluster nodes will never reach the base station.

In our protocol, besides having a CH in the cluster, there is a sub-CH that takes the role of the CH when the CH dies because the reasons we mentioned above.

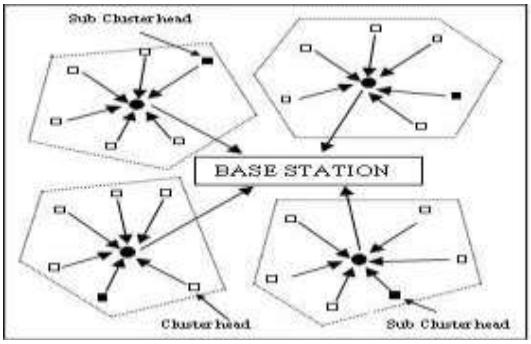

Figure 3: LEACH-sub-CH Protocol Architecture.

V. Conclusion

Data mining, clustering and data aggregation in clustered multi feature sensor network is very important and difficult task. To overcome the drawbacks of the LEACH protocol, this paper proposes an improved version of the well known protocol for wireless sensor networks called *LEACH protocol* which is the first and the most important protocol in wireless sensor network which uses cluster based broadcasting technique. This improved or a new version of LEACH protocol called as Improved LEACH or LEACH-sub-CH protocol. Using this protocol we are building a method called *MFILC*, which is used to cluster the sensor nodes in the network.

Therefore, the new version of LEACH outperforms the original version of LEACH protocol. As another improvement, the cluster heads could form a multi-hop backbone whereby data are transmitted among cluster heads until they reach the BS. Alternatively, LEACH can evolve into a hierarchical protocol by forming *"super-clusters"* out of the cluster head nodes and having a *"super-cluster head"* that processes the data from all the cluster head nodes in the super cluster. These changes will make LEACH suitable for a wider range of wireless sensor networks.

References

[1] Purnima Khuntia, Prasant Kumar Pattnaik "Some Target Coverage Issues of Wireless Sensor Network", International Journal of Instrumentation, Control & Automation (IJICA), Volume 1, Issue 1, 2011

[2] Ian F.Akyildiz, WeilianSu, Yogesh Sankarasubramanian, and Erdal Cayirci Georgia Institute of Technology "A Survey on Sensor Networks" IEEE Communication Magazine August 2002.

[3] Yong Wang, Garhan Attebury and Byrav Ramamurthy "A Survey Of Security Issues In Wireless Sensor Networks" IEEE Communications surveys, The Electrinoc Magazine of Original Peer-Reviewed Survey Articles, Volume 8, No. 2, 2006.

[4] Hyuk Park, Soo Young Moon, Tae Ho Cho, "A Region Segmentation Based Path Selection Method for WSNs", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.2, February 2011

[5] S. Guinepain and L. Gruenwald, "Automatic Database Clustering Using Data Mining," Database and Expert Systems Applications, DEXA '06. 17th International Conference on , pp.124-128, 2006

[6] J. Yu and P. Chong, "A Survey of Clustering Schemes for Mobile Ad Hoc Networks", IEEE Communications Surveys & Tutorials, 2005.

[7] T. Chan, C. Chen, Y. Huang, J. Lin and T. Chen, "Optimal Cluster Number Selection in Ad-hoc Wireless Sensor Network," WSEASTRANSACTIONS on COMMUNICATIONS Manuscript received Jan 12, 2008; revised July 10, 2008

[8] F. Xiangning, and S. Yulin, "Improvement on LEACH Protocol of Wireless Sensor Network," In Proceedings of the 2007 international Conference on Sensor Technologies and Applications SENSORCOMM, IEEE Computer Society, Washington, pp. 260-264, 2007.

[9] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy- Efficient Communication Protocols for Wireless Microsensor Networks," in International Conference on Systems Science, 2000.

[10] O. Younis , S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, "Mobile Computing, IEEE Transactions on, Vol. 3, No. 4, pp. 366-379, 2004.

[11] A. Taherkordi, R. Mohammadi,and F. Eliassen, „A Communication- Efficient Distributed Clustering Algorithm for Sensor Networks," In Proceedings of the 22nd international Conference on Advanced information Networking and Applications - Workshops , 2008.

[12] M. Xiuli , L. Shuangfeng , L. Qiong , Y. Dongqing and T. Shiwei, "Distributed, Hierarchical Clustering and Summarization in Sensor Networks, " Advances in Data and Web Management book chpter, ISBN 978-3-540-72483-4, 2007.

[13] Rabie A. Ramadan, "Data Mining in Multi-Feature Sensor Networks", IEEE, 2009.

[14] Y.C. Wang and Y.C. Tseng, "Intentional Mobility in Wireless Sensor Networks," Wireless Networks: Research, Technology and Applications, Nova Science Publishers, 2009.