Performance Analysis of LEACH with Machine Learning Algorithms in Wireless Sensor Networks

Sukhchandan Randhawa Computer Science and Engineering Department Thapar University, Patiala, India

ABSTRACT

Wireless Sensor Networks consist of thousands of power constrained micro sensors whose main task is to sense and report the target phenomena to the base station. Hierarchical routing plays an important role for transmitting the aggregated data to the sink. Sensor nodes are organized into number of clusters and within each cluster, cluster head is responsible for collecting the data and to report that data to the Base Station. Machine learning algorithms play an important role while selecting the cluster head based on various QoS parameters. In this paper, a hierarchical protocol LEACH is chosen for analyzing the impact of machine learning algorithms - K-Means and modified K-Means clustering on energy consumption of nodes by varying the type of input parameters. This paper covers the brief introduction of 802.15.4 based Wireless Sensor Networks, power models, machine learning algorithms for sensor clustering and simulation environment using NetSim.

General Terms

Networks, Wireless Sensor Networks, Routing Protocols, Energy Efficiency.

Keywords

Energy Efficiency, Hierarchical Routing, Clustering, Wireless Sensor Networks.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) plays an important role in number of application like health monitoring, smart cities, whether forecasting and wildlife monitoring etc[1]. A WSN consists of large number of small sized, power constrained sensor nodes which have limited communication and computation abilities. These tiny nodes sense the target phenomena and send the gathered data to the sink node. Routing plays a vital role in communication. Hierarchical routing is an efficient mechanism to send the data from source to destination [2]. In hierarchical routing, an entire region is divided into number of clusters and within each cluster, a cluster head is elected based on some predefined parameter [3]. The parameters which are selected while making cluster head selection have a huge effect on energy level. Communication networks have become too complex for traditional analytical approaches or "rules of thumb" methods to provide an accurate understanding of system behavior and possible problems and solutions. So based on machine learning clustering approaches, the behavior of Low Energy Adaptive Clustering Hierarchy (LEACH) protocol has been analyzed in terms of energy efficiency and network lifetime. There are two clustering algorithms which are used for cluster head selection in LEACH protocol: K-Means clustering protocol which considers distance as a sole parameter and Modified K-Means clustering protocol which considers

Sushma Jain Computer Science and Engineering Department Thapar University, Patiala, India

residual energy and distance both as input parameters for cluster head selection in LEACH protocol.

This paper is organized as follows: Section 2 introduces Low Power IEEE 802.15.4 based Wireless Sensor Networks and Power Models which can reduce the energy consumption. Section 3 gives the brief overview of clustering along with Kmeans clustering technique. Section 4 shows the simulation results and simulation platform being used for analysis. Finally, in Section 5, the conclusions are presented.

2. LOW POWER IEEE 802.15.4 BASED WIRELESS SENSOR NETWORKS

IEEE 802.15.4 standard [4][5] specifies Physical and Media Access Control layer for low rate personal area networks. This standard focuses on very low cost low power communication between devices with little or no underlying infrastructure. This is in complete contrast to Wi-Fi which offers more bandwidth and requires more power. The basic communication range is of 10m approximately with transfer rate of 250 kbps. CSMA/CA is used for collision avoidance while data transmission. The operating frequencies are in range of 868-915 MHz with 2.4 GHz ISM band which is free of cost available.

2.1 IEEE 802.15.4 Power Models

WSN has two operational modes namely *Beacon Enabled* and *Non Beacon Enabled mode* as shown in Fig. 1. In Beacon Enabled mode, nodes follow Superframe structure which is composed of two parts: Contention Access Period (CAP) without Guaranteed Time Slots (GTS) and Contention Free Period (CFP) with GTS [6].





A coordinator can optionally bound its channel time by using a superframe structure as shown in Fig. 2. A superframe is bounded by transmission of a beacon frame and can have an active and an inactive portion. The coordinator may operate in a low power sleep mode during an Inactive portion. The active part is divided into sixteen equally spaced slots and is composed of three parts: *a beacon, CAP* and *CFP*. Beacons are periodically generated by the coordinator. These are used to synchronize the attached devices and to describe the structure to the superframe. In contention access period, all nodes have a packet to transmit, use Slotted CSMA /CA mechanism to access the channel. CFP is used to allocate slots for the nodes which require low latency. For this purpose, coordinator allocated GTS for such nodes.



Fig. 2: Superframe Structure

3. LEACH PROTOCOL

LEACH is a hierarchical procedure in which the nodes transmit their data to the CHs, and these CHs are cooperative in nature which condense the data and forward it to the Base Station (BS).

At network layer, Dynamic Source Routing (DSR) is used and at MAC layer Zigbee protocol i.e. 802.15.4. Stochastic algorithm is used by each node in each round for resolving access problems. LEACH is a protocol that is used for dynamic clustering in WSNs. This protocol randomly selects the cluster heads in a network so that energy can be distributed evenly among the among the sensor nodes [7]. Data collected by cluster heads from sensor nodes is transmitted to the base station. This is needed because a sensor node is not of any use if its battery dies. Each time when an updation process of cluster restructuring is initiated, then it is known as a round and further each round is subdivided into two stages: set-up phase and steady phase as shown in Fig. 3. The set-up stage includes three sub-stages called advertisement stage, cluster development and schedule development.

In first phase, CH of each cluster broadcast its identity to the sensor nodes whereas in steady phase CH gets all the information given by sensor nodes. CHs consume large amount of energy as compared to cluster member nodes in communication. Therefore, to cop up with this issue, LEACH proposed that every member of cluster should get equal chance to become CH so, that energy dissipation can be balanced in a network.

In this, random value between 0 and 1 is assigned to each node in each round. That particular node becomes the CH if its value is less than the threshold value [8].



Fig. 3: Formation for Cluster Round in LEACH protocol.

Therefore, by this method energy consumption in a network can be balanced by giving all nodes a fair chance to elect as CH. But this approach does not provide information about node location during the selection of CHs. As a result, this protocol may lead to the unbalanced energy in the network.

$$T(n) = \frac{P}{1 - P * \left(r \mod \frac{1}{P}\right)} \quad \forall n \in G$$
$$= 0 \quad \forall n \notin G \quad (1)$$

In eq. (1), P refers to the percentage of number of CHs in a network, r belongs to the present round number, and G belongs to the set of sensor nodes which are not being selected as CH in the past 1/P rounds. A random value between 0 and 1 is given to each node in each round. If value is lower than the threshold value T(n), then that node is elected as the CH. Sometimes nodes that are CHs in present round don't get chance to become CH once more for the next P rounds and to become CH once more, each node has 1/P probability. Lastly, in each round, non-cluster head nodes chose the closest cluster head to join their clusters. After this, CH creates a schedule for each node in its cluster to transmit its data.

All non-cluster head nodes mainly converse with the CH nodes in a Time Division Multiple Access (TDMA) fashion based on the program given by CHs. This can be achieved if less amount of energy is needed by sensor nodes to reach the CH. Code Division Multiple Access (CDMA) is also used by LEACH so that abnormal set of CDMA codes are used by each cluster in order to reduce hindrance among clusters in the network. In this protocol, nodes get self-sufficient opinion for the formation of clusters by means of a dispersed algorithm of any centralized control. By knowing the correct location of the sensor nodes, outline of a cluster can be done in the WSN and also for setting up the clusters, no global communication is needed. It is assumed that all the nodes start with the same amount of energy and the cluster model should be intended in a way that nodes are CHs[9]. The space required for the noncluster head nodes to send data can be reduced by extending the CH nodes in overall network. After the node has become CH, it broadcasts an announcement message which is small in size and contains the node's ID and a header that differentiate this message as an affirmation message. From this announcement message CH signal strength can be estimated by which non-cluster head nodes decide their CHs to reside in their clusters. After the selection of clusters and their CHs by each node, it should report to the CH node that it is a member node of its cluster now by transmitting a join-request message (JoinREQ) back to the chosen CH. In LEACH, synchronization of the data conveyance can be done in the clusters by the CHs which act as local control centers. TDMA agenda is used by CH nodes and the communication of this program is done with the nodes in the cluster. It is assured that there is zero collision while transmitting the data messages and also the radio equipment of each non CH node is turned off all the time except the transmitting time. Therefore, it leads to the minimization of energy dissipation by the individual node.

4. CLUSTERING TECHNIQUES

Clustering is considered as the most important *unsupervised learning* problem. The main objective of clustering is to find a pattern in a collection of unlabeled data [10]. Clustering can be defined as "the process of organizing objects into groups whose members are similar in some way". A *cluster* is defined as a collection of data items which have similarity

among them and are dissimilar to the data points belonging to other clusters. The data is divided into clusters based on some criterion which is usually a *distance*: two or more data points belong to the same cluster if they are close according to a given distance. This type of clustering is known as *distancebased clustering*. Another type of clustering is *conceptual clustering* in which two or more objects belong to the same cluster if a concept is defined which is *common* to all that objects. There is no absolute "best" criterion which is independent of the final objective of the clustering. Consequently, it is responsibility of the user to decide this criterion, in such a way that the result of the clustering is according to the requirements.

The clustering algorithms may be classified as shown [13] in Fig. 5:



Fig. 5: Types of Clustering Algorithms

In the *Exclusive Clustering*, data are grouped in such a way, so that if a certain data item belongs to one cluster then it cannot be included in another cluster. A simple example of that is shown in Fig. 6, where a straight line provides a clear separation of points on a two-dimensional plane. On the contrary in overlapping clustering, fuzzy sets are used to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, an appropriate membership value will be associated to data values.



Fig. 5: Exclusive Clustering

A hierarchical clustering algorithm is calculated on the basis of the union between the two nearest clusters. The initial condition is defined by taking each data value as a cluster. After a few iterations it reaches the final clusters wanted. Finally, the last type of clustering uses a probabilistic approach. K-means is an exclusive clustering algorithm; Fuzzy C-means is an overlapping *clustering* algorithm. In this research paper, K-means clustering algorithm and Modified, K-means clustering algorithm have been used which are discussed in the following sub-sections.

4.1 K-Means Clustering algorithm

K-means is the simplest unsupervised learning algorithm that solves the well known clustering problem. A simple and easy way is followed to classify a given data set into a certain number of clusters (k clusters) which are prefixed [11][12]. The main idea is to define k centroids, one for each cluster.

These centroids should be placed in an optimal way. So, the better choice is to place these centroids as far as possible away from each other. In the next step, each point is taken belonging to a given data set and is associated to the nearest centroid. When all the points are mapped to some cluster, then the initial step is completed and an initial grouping is done. At this point, again k new centroids are calculated as barycenters (which is the centre of mass of two or more bodies) of the clusters resulting from the previous step. After k new centroids are formed, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been executed. Due to execution of this loop, it may be noticed that the k centroids change their positions step by step until there are no more changes need to be accomplished. In other words centroids do not move any more. This algorithm targets at minimizing an objective function, which is usually a squared error function. The objective function is defined in (eq. 2)

$$J = \sum_{j=l}^{k} \sum_{i=1}^{x} \|X_{i}^{(j)} - C_{j}\|^{2}$$
(2)

where $||X_i^{(j)} - C_j||^2$ is a chosen distance measure between a data point $X_i^{(j)}$ and the cluster centre C_j , is an indicator of the distance of the *n* data points from their respective cluster centroids. The algorithm is composed of the number of steps as shown in Fig. 6:

Algorithm: K-Means Clustering algorithm

- 1. Place *K* points into the space represented by the objects that are being clustered. These points represent initial group centroids.
- 2. Assign each object to the group that has the closest centroid.
- 3. When all objects have been assigned, recalculate the positions of the *K centroids*.
- 4. *Repeat Steps 2 and 3 until* the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Fig. 6: Pseudo code of K-Means clustering algorithm

The procedure will always terminate but, it does not necessarily choose the most optimal configuration, corresponding to the global objective function minimum. The algorithm is significantly sensitive to the initial randomly selected cluster centroids. So, this algorithm can be run multiple times to reduce this effect.

5. SIMULATION ENVIRONMENT

NetSim has been interfaced with MATLAB [14]. MATLAB is used to carry out the complex mathematical computations for WSNs making coding more elegant and easy to understand. NetSim uses MATLAB for computing its cluster distribution [15]. For WSN simulations, NetSim models the sensor network, initializes the various protocols for communication, senses the agent, implements mobility, determines the route to the sink, transmits information among the sensors, does packet tracing, handles events and generates the output statistics. The MATLAB interface allows NetSim to pass instantaneous network data which includes sensors' positions and power. MATLAB then uses its mathematical tools to compute the clustering and then sends back the clustering data to NetSim. NetSim then uses this data to carry out its network simulation routines as shown in Fig. 7.

One of the aspects of machine learning is to find out patterns in data. Three special mathematical techniques for finding



Fig. 8: Formation of Clusters and cluster heads in LEACH protocol

K-Means clustering algorithm is the first clustering algorithm which is applied. K-Means algorithm partitions data into k number of mutually exclusive clusters. The MATLAB function for K-Means based clustering is defined as follows:

$$[IDX,C] = kmeans(X,k)$$
(3)

where X is the matrix containing x, y coordinates of the sensors, k is the number of clusters, *IDX* contains the cluster id's of each sensor i.e. the cluster to which the sensor belongs and C represents the centroid of each cluster.





5.1 K-Means Clustering algorithm

In this Clustering method, k-Means clustering method is used in which only parameter i.e. distance is considered. 100 sensors are placed evenly on XY plane and each sensor start consuming energy and z- axis represents the power consumed. As it can be seen from plot, there are four peaks in the plot corresponding to the four sensors which have been selected as Cluster Heads (CHs) As shown in Fig. 9. K-means clustering algorithm selects CHs for each of the clusters and then all the data of the clusters is transmitted to the CHs. then these CHs send their data to the base station. Since the cluster head transmits and receives the more data as compared to the rest of the sensors, so the energy consumed by the CHs is far higher than the energy consumed by the other sensor.



Fig. 9: Energy consumption of CHs in K-Means Clustering algorithm

Simulation results of this approach are shown in the Fig. Thus the K-Means algorithm leads to highly uneven energy consumption. This leads to cluster heads running out of power very quickly and hence leads to a very short network lifetime. The energy consumed by these CHs is four times the energy consumed by the other sensor nodes.

5.2 Modified K-means Clustering algorithm

Modified K-means Clustering algorithm use distance and residual energy as parameters as compared to the previously discussed K-Means algorithm in which distance is taken as sole parameter for decisions. The use of energy as input parameter ensures that not just distance but residual energy also used to determine the cluster heads.

In the initial phase, the plot resembles the previous one, however as the time passes, it can be observed that the power consumed by all the sensors is approximately at the same rate as shown in Fig. 10. There are no peaks in the plot unlike the previous one, because modified K-Means algorithm takes into account the energy level of each sensor and thus different sensors will be appointed as CHs.

This approach leads to the uniform power consumption in the network and thus allows the network lifetime significantly longer when compared to prior K-means algorithm. The power consumption in CHs in modified K-Means approach is 480 to 500 mW. In prior K-means approach, the power consumption is 700mW which is four times the normal power consumption



Fig.10: Energy consumption of CHs in Modified K-Means Clustering algorithm

6. CONCLUSION AND FUTURE SCOPE

IEEE 802.15.4 based WSNs are low powered networks. Hierarchical Routing (Cluster based routing) plays an important role while transmitting the data from source to the destination, as it is one of the energy efficient techniques. The Cluster Heads are elected in these protocols based on some parameters. The selection of these input parameters plays an important role. In this research paper, to analyze the impact of clustering techniques and their input parameters, LEACH (hierarchical protocol) has been selected for analysis. In the first case, K-Means Clustering algorithm along with distance as the only parameter. Modified K-Means Clustering algorithm which takes into account two parameters- distance as well as residual energy is also considered. As evident from results, the later approach performs better in terms of Residual energy of CHs and network lifetime as compared to the prior approach. The energy consumed in the first approach is almost 4 times more than the energy consumption by normal nodes. In future, more clustering techniques like Fuzzy-C means, hierarchical clustering etc. can be implemented for choosing the best suitable approach for the underlying routing protocol. Mobility of nodes can also be considered for making the simulation environment more dynamic.

7. REFERENCES

- Akyildiz I.F., Su W., Sankarasubramaniam Y. and Cayirci E., 2002 "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, pp. 393–422.
- [2] Hosseinzadeh M. and Alguliev R. M. 2010, "Hierarchical routing in wireless sensor networks: a survey," 2nd International Conference Computer Engineering Technology, vol. 3, pp. V3–650–V3–654.
- [3] Iwanicki K. and Steen M. V., 2009 "On hierarchical routing in wireless sensor networks," Ipsn'09, pp. 133– 144.
- [4] Yu Q., Xing J. and Zhou Y., 2006 "Performance Research of the IEEE 802.15.4 Protocol in Wireless Sensor Networks," 2nd IEEE/ASME International Conference on Mechatronics and Embedded Systems and Applications, pp. 1 – 4.
- [5] Lu J., Bossche A. V. D. and Campo E. 2014 "An IEEE 802.15.4 Based Adaptive Communication Protocol in Wireless Sensor Network: Application to Monitoring the Elderly at Home", Wireless Sensor Network, Vol. 6, pp. 192–204.
- [6] Mura M., Paolieri M., Fabbri F., Negri L. and Sami M. G., 2007 "Power Modeling and Power Analysis for IEEE 802.15.4: a Concurrent State Machine Approach", 4th IEEE Conference on Consumer, Communications and Networking, pp. 660 – 664.
- [7] Malik M., Singh Y. and Arora A., 2013, "Analysis of LEACH Protocol in Wireless Sensor Networks," vol. 3, no. 2, pp. 178–183.
- [8] Fu C., Jiang Z., Wei W., and Wei A., 2013, "An Energy Balanced Algorithm of LEACH Protocol in WSN," International Journal of Computer Science, vol. 10, no. 1, pp. 354–359.
- [9] Thein M. C. M. and Thein T., 2010 "An energy efficient cluster-head selection for wireless sensor networks," ISMS - UKSim/AMSS 1st International Conference on Intelligent Systems, Modeling and Simulation, pp. 287– 291.

International Journal of Computer Applications (0975 – 8887) Volume 147 – No.2, August 2016

- [10] Xu R., 2005 "Survey of clustering algorithms for MANET," IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645–678.
- [11] Alsabti K., Ranka S., and Singh V., 1997, "An efficient k-means clustering algorithm," Electrical Engineering and Computer Sciences.
- [12] Kanungo T., Mount D. M., Netanyahu N. S., Piatko C. D., Silverman R., and Wu Y., 2002 "An efficient kmeans clustering algorithm: analysis and implementation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881–892.
- [13] http://home.deib.polimi.it/matteucc/Clustering/tutorial_ht ml accessed on 15th June 2016.
- [14] Abo-zahhad M., Amin O., Farrag M. and Ali A., 2014 "A Survey on Protocols, Platforms and Simulation Tools for Wireless Sensor Networks," International Journal of Energy, Information and Communication., vol. 5, no. 6, pp. 17–34.
- [15] http://tetcos.com/blog/2016/02/17/interfacing-netsimwith-matlab/ accessed on 10th June 2016.