Improved Load Balanced Clustering Algorithm for Wireless Sensor Networks

Pratyay Kuila and Prasanta K. Jana

Department of Computer Science and Engineering
Indian School of Mines, Dhanbad-826 004, India
{pratyay_kuila,prasantajana}@yahoo.com

Abstract. In this paper, we present an improved load balanced clustering scheme for wireless sensor networks. We show that the algorithm runs in $O(n \log n)$ time for n number of sensor nodes with a simpler problem in which all the sensor nodes have equal loads. We prove that the algorithm is optimal in assigning sensor nodes to the available gateways. We also show that the same scheme is also applicable for the scenario in which the sensor nodes may have unequal loads. We prove that the scheme for the later is a 2-approximation of load balanced clustering algorithm.

Keywords: Wireless sensor networks, clustering, load balancing, approximation algorithm, network lifetime.

1 Introduction

The wireless sensor networks now witness the increased interest in the potential use in various application areas such as disaster management, military, environment, health, home and disaster relief [1]. A WSN is composed of a large number of sensor nodes which are randomly or manually deployed in a coverage area. All the sensor nodes collect local information, process them and send it to a remote base station (called sink). The sink is connected to the Internet for the public notice of the phenomena. One of the most important constraints on sensor nodes is the requirement of low power consumption. Reducing energy consumption to maximize the network lifetime is thus considered as the most critical challenge in WSN. Many research articles [2], [3] have been addressed on this issue. However, design of energy efficient clustering algorithms is the most promising area in this regard.

In many applications of the WSN, some high-energy nodes called gateways are typically deployed along with the sensor nodes. Sensor nodes are then efficiently grouped into distinct clusters with a single gateway for each cluster. Each sensor node belongs to only one cluster and each gateway is treated as the cluster-head (CH) of a cluster. The sensor nodes inside a cluster can communicate with the CH either through single hop or multi-hop communication. Therefore, each sensor node has certain processing and communication load, which is jointly termed as the load of the sensor node. The gateways collect and process the local data and send the processed data to the sink by multi-hop communication through other gateways. However,

P.S. Thilagam et al. (Eds.): ADCONS 2011, LNCS 7135, pp. 399–404, 2012.

improper assignment of the sensor nodes can make some gateways overloaded with high number of sensor nodes. This can reduce the network life time and degrade the overall performance of the WSN. Therefore, we are concerned here of assigning the sensor nodes to the gateways (for the formation of clusters) such that the maximum load of each gateway is minimized. This problem is actually referred as the load balanced clustering problem (LBCP) in the literature [4].

A number of clustering algorithms [5], [6], [7] have been developed for WSN. LEACH [8] is a popular clustering technique that forms clusters by using a distributed algorithm. However, the main disadvantage of this approach is that a node with very low energy may be randomly selected as a cluster head which can die quickly. Therefore, a large number of algorithms have been developed to improve LEACH such as PEGASIS [9], HEED [10], TEEN [11] and TL-LEACH [12]. Other clustering algorithms have also been recently reported in [13], [14], [15].

Low et al. [4] proposed a load balanced clustering algorithm for WSN that has the time complexity of $O(mn^2)$ assuming equal loads of the sensor nodes. It is shown that LBCP with unequal load for sensor nodes is NP-hard [4].

In this paper, we develop an improved load balanced clustering algorithm that requires $O(n \log n)$ time in contrast to $O(mn^2)$ time required by Low [4]. We also show that this is a 2-approximation algorithm for unequal loads of the sensor nodes.

The paper is organized as follows. Section 2 presents the required terminologies. Section 3 discusses the proposed algorithm followed by the conclusion.

2 Terminologies

Here, we consider a WSN scenario with the same assumption as that of [4] that is, gateway nodes are chosen a priori and are fixed throughout the network lifetime. The gateways are less constraint than the sensor nodes. A sensor can be assigned to any gateway if it is within its communication range. Therefore, there are some pre-specified gateways onto which a particular sensor may be assigned. So, each sensor has a gateway list and this sensor can be assigned to exact one gateway from this list. Even, the probability of the assignment to a particular gateway from the list is different from other gateways. We adopt the following assumptions and notations similar to [4].

1) The set of sensor nodes is denoted by $S = \{s_1, s_2, \cdots, s_n\}$.
2) The set of gateways is denoted by $G = \{g_1, g_2, \cdots g_m\}$.
3) $n > m$, i.e. the number of sensor nodes is greater than the number of gateways.
4) The load contributed by each sensor node can be estimated.
5) d_i denotes the traffic load contributed by sensor node s_i where $s_i \in S$ and $d_i \in N$, N is the set of natural number.

6) G_i denotes the set of gateways to which sensor s_i may be assigned, where $s_i \in S$ and $G_i \subseteq G$. We note that some constraints may be imposed such that a given sensor s_i can only be assigned to a member of a selected set of gateways G_i

with some probability. Let there be a sensor s_k and its possible gateway list be $G_k = \{g_1, g_2..., g_r\}$ onto which s_k can be assigned. Let, $P_i(x)$ be the probability of assigning s_i to the gateway g_x. Then by the law of probability, we have

$$\sum_{x=1}^{m} P_i(x) = 1 \mid \forall s_i \in S, g_x \in G \tag{2.1}$$

7) The average possible load (APL) of a gateway g_x is the summation of mean loads of all the sensor nodes with the probability that they can be assigned to g_x. Therefore the APL can be expressed as follow.

$$APL(g_x) = \sum_{i=1}^{n} \{P_i(x) \times d_i \mid \forall s_i \in S, g_x \in G_i \tag{2.2}$$

If any sensor node, say s_i is finally assigned to the gateway say g_x .Then the APL of the gateway g_x and the APL of any other gateway, say g_y can be updated as follows.

$$APL(g_x) = APL(g_x) + (1 - P_i(x)) \times d_i \tag{2.3}$$

$$APL(g_y) = APL(g_y) - P_i(y) \times d_i, \forall g_y \in G_i - \{g_x\} \tag{2.4}$$

8) Let W_i be the load assigned to the gateway g_i. Then the overall maximum load of the gateways is

$$W = \max\{W_i \mid \forall g_i \in G\} \tag{2.5}$$

Our main objective is to design a load balanced clustering algorithm which can minimize the overall maximum load of the gateways i.e., W.

3 Load Balanced Clustering Problem with Equal Load

We consider here a special case of load balanced clustering problem, where each sensor node has equal load, i.e., $d_j = \alpha$ (say), for some constant α, $\forall s_j \in S$, $1 \leq j \leq n$. So, minimizing the overall maximum load of the gateway is equivalent to minimizing the maximum number of sensor nodes that can be assigned to each gateway.

The basic idea of our algorithm is as follows. We first sort all the sensors S in ascending order on the number of gateways to which sensor s_i may be assigned, $1 \leq i \leq n$. Let, $S = \{s_a, s_b, s_c, \ldots, s_p\}$ be this sorted sensor list. We now, successively consider these sorted sensor nodes starting with s_a for their assignment to the correct gateway. In order to assign s_a, we consult its corresponding set of possible gateway i.e., G_a and calculate the APL values using equation (2.2) for all the gateways belongs to G_a and assign s_a to that gateway that has the minimum APL. If there are two or more gateways with the same APL then, select that gateway with highest probability that contributes most to the APL value. If probability is also same then select that gateway with minimum number of sensors already assigned to it. After each assignment of sensor, the APL of the gateways are updated by equations (2.3) and (2.4) for the assignment of the next sensor from the sorted list. The same procedure is continued until all the sensor nodes are allotted to their correct gateway.

Algorithm: LBClustering

Input:

1. A set of sensors $S = \{s_1, s_2..., s_n\}$ with same load $d_i = \alpha,\ 1 \le i \le n$.
2. A set of gateways $G= \{g_1, g_2..., g_m\}$.
3. For each s_i, the set of gateways G_i and the probability $P_i(x)$, $\forall g_x \in G_i$.

Output: An assignment $A : S \rightarrow G$ such that W is minimized.

Step 1: Sort the sensor nodes in ascending order of number of possible gateways.
Step 2: For j = 1 to n
{
While ($G_j \neq \Phi$) /* Φ = NULL */
$APL[k] := APL[k] + P_j(k).d_j,\ \forall g_k \in G_j$
}
Step 3: For j = 1 to n /* Assignment of sensor nodes to gateways */
{
3.1: While ($G_j \neq \Phi$)
{
Select the gateway with minimum *APL* value calculated in Step 2.
If there is two or more gateway with same *APL* value, then select that gateway with highest probability, which contributes most to the *APL* value. If probability is also same then select that gateway with minimum number of sensors already assigned.
}
3.2: Let, g_k be the selected gateway from G_j.
3.3: Assign s_j to g_k;
3.4: $APL[k] := APL[k] + (1 - P_j(k)).d_j$;
3.5: While ($G_j \neq \Phi$)
$APL[x] := APL[x] - P_j(x).d_j,\ \forall g_x \in G_j - \{g_k\}$
}
Step 4: Stop

Theorem 1: The LBClustering produces optimal solution in $O(n \log n)$ time assuming equal loads of the sensor nodes.

Proof: Let $S = \{s_1, s_2, s_3,..., s_n\}$ be the sorted list after step 1. For the successive assignment of $s_1, s_2, s_3, s_4,...,s_n$, the following approach is followed for load balancing. The algorithm assigns that sensor node first, which has the least chance of assigning to a gateway by the steps 2 and 3. As a result any other sensor node having more chance can select a gateway with the least load. Let s_k be the last assigned sensor node to the maximum loaded gateway g_r, after its final assignment. Let current load on g_r be l_r. Suppose that there is a gateway $g_s \in G_k$ with total load l_s such that $l_s < l_r$. Therefore, s_k can be assigned to g_s and current load of g_s will be $l_s + \alpha$, where α is the load of each sensor nodes. Then the algorithm is not optimal if $(l_s + \alpha) < l_r$. But at the time of assignment of s_k, the selected gateway g_r had the minimum *APL* i.e., $(l_r - \alpha)$ was minimum. So, after assigning of s_k to any $g_s \in G_k - \{g_r\}$, $(l_s + \alpha) \geq l_r$. Hence its is optimal.

For the time complexity, we proceed as follows. Step 1 requires $O(n \log n)$ time for sorting. As the while loop inside the step 2 can take at most $O(m)$ time, step 2 is executed $O(mn)$ time. Steps 3.1 to 3.5 can be done in $O(m)$. Outer for loop of step 3 requires $O(n)$ time. So, step 3 can be executed in $O(mn)$ time. Therefore the above algorithm requires $O(mn)$ time. If $m < \log n$, it requires $O(n \log n)$ time.

Theorem 2: The Load-Balanced-Clustering algorithm is a 2-Approximation algorithm of LBCP with unequal load of the sensor nodes.

Proof: Let *OPT* be the maximum load of a gateway in an optimal solution. Then it is easy to see that $OPT \geq d_i \ \forall i,\ 1 \leq i \leq n$. Let I be the smallest instance of the problem for which this algorithm conflicts with optimal solution. Let, g_i be a gateway with maximum load after complete run of the Load-Balanced-Clustering algorithm, i.e., $W_i = max\{W_j \mid \forall j,\ 1 \leq j \leq n\}$. Let s_r be the last sensor node assigned to g_i. The crucial property of our algorithm is that, at the time of the assignment of s_r, g_i was the minimum loaded gateway from G_r. So, before assignment of s_r, load of g_i was $(W_i - d_r)$, which is less than or equal to $OPT(I)$. So, $W_i - d_r \leq OPT(I)$. As, $OPT \geq d_i \ \forall i,\ 1 \leq i \leq n$, $W_i \leq 2OPT(I)$. Therefore this is never more than a factor 2 from optimal solution.

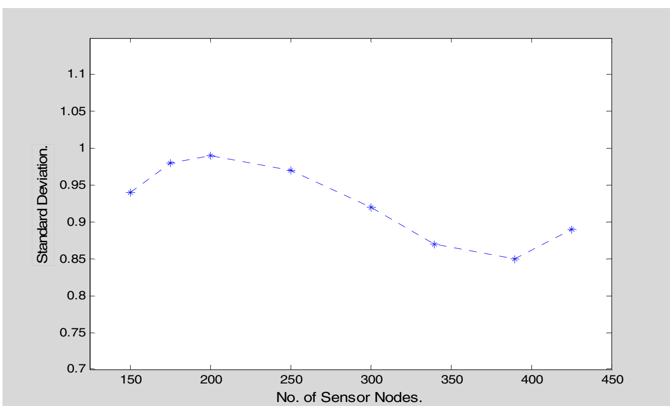

Fig. 1. Plot of standard deviation vs. number of sensor nodes

Experimental Result: The proposed algorithm was run on an environment with 30 gateways by varying the number of sensor nodes from 150 to 425 and varying the load of the sensor nodes from 2 to 6. The standard deviation of the loads of the gateways is calculated and plotted against the number of sensor nodes as shown in Fig. 1. It can be observed that, the value of the standard deviation varies within the range 0.85 to 1.0 which shows the efficiency of the algorithm. It starts rising from 150 number of nodes and decaying from 200 and it is least at 390.

4 Conclusion

An improved load balanced clustering algorithm has been presented that runs $O(n \log n)$ time. It has been shown to be optimal in assigning sensor nodes to the gateways for equal loads of the sensor nodes. The same scheme has been shown to work as a 2-approximation algorithm for unequal loads of the sensor nodes. Our future attempt will be made to device improved approximation algorithm for unequal loads.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless Sensor Networks: A Survey. Computer Networks 38(4), 393–422 (2002)
2. Kim, K.T., et al.: An Energy Efficient Routing Protocol in Wireless Sensor Networks. In: Proc. of Int. Conf. on Comp. Sc. and Engg., pp. 132–139. IEEE (2009)
3. Lattanzi, E., et al.: Energetic sustainability of routing algorithms for energy-harvesting wireless sensor networks. Computer Comm. 30, 2976–2986 (2007)
4. Low, C.P., et al.: Efficient Load-Balanced Clustering Algorithms for wireless. Computer Communications 31, 750–759 (2008)
5. Abbasi, A.H., Younis, M.: A Survey on clustering algorithms for wireless sensor networks. Computer Communications 30, 2826–2841 (2007)
6. Boyinbode, O., et al.: A Survey on Clustering Algorithms for Wireless Sensor Networks. In: 13th Intl. Conf. on Network-Based Information Systems, pp. 358–364. IEEE (2010)
7. Jiang, C., Yuan, D., Zhao, Y.: Towards Clustering Algorithms in Wireless Sensor Networks-A Survey. In: Wireless Communications and Networking Conference (WCNC), pp. 1–6. IEEE (2009)
8. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: Application specific protocol arch. IEEE Tran. on Wireless Comm.s. 1(4), 660–670 (2002)
9. Lindsey, S., Raghavendra, C.S.: PEGASIS: power efficient gathering in sensor information systems. In: Proc. of the IEEE Aerospace Conference, vol. 3, pp. 1125–1130 (2003)
10. Younis, O., Fahmy, S.: HEED: A Hybrid, Energy-Efficient, Distributed clustering approach for Ad Hoc sensor networks. IEEE Trans. 3, 366–379 (2004)
11. Manjeshwar, A., Agrawal, D.P.: TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In: Proc. of the 15th Intl. Workshop on Parallel and Distri. Comp. Issues in Wireless Networks and Mobile Computing, San Francisco, pp. 2009–2015 (2002)
12. Loscri, V., et al.: A Two-Level Hierarchy for Low-Energy Adaptive Clustering Hierarchy. In: Proc. of Vehicular Technology Conference, vol. 3, pp. 1809–1813. IEEE (2005)
13. Liu, Z., Zheng, Q., Xue, L., Guan, X.: A distributed energy-efficient clustering algorithm with improved coverage in wireless sensor networks. Future Generation Computer Systems, 1–11 (2011)
14. Min, X., Ren, S.W., Chang-jiang, J., Ying, Z.: Energy efficient clustering algorithm for maximizing lifetime of wireless sensor networks. Int. J. Electronics Communication. (AEÜ) 64, 289–298 (2010)
15. Dimokas, N., Katsaros, D., Manolopoulos, Y.: Energy-efficient distributed clustering in wireless sensor networks. J. Parallel Distributed Computing 70, 371–383 (2010)