International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

A Novel Method for Data Aggregation using Spatial Correlation Clustering Method

Savita Patil[†*] and Sunita Rawat[†]

[†]Dept of ISE, The Oxford College of Engineering, Bangalore, India

Accepted 20 June 2015, Available online 25 June 2015, **Vol.5, No.3 (June 2015)**

Abstract

In Wireless Sensor Networks, the sensor nodes are the devices that are capable of collecting, sensing and gathering data from environment specified by the user. In order to save limited resources of sensor nodes, it is required to aggregate data from the nodes. In data aggregation process it aggregates all sensor nodes data with least amount energy utilization and sends its data to the destination. The main goal of data aggregation algorithms is to gather and collects data in an energy efficient manner so that network lifetime is improved. When sensor nodes are densely deployed, collecting data from nodes is a major task and to simplify this process, many data aggregation methods are studied and one of the data aggregation method is sending local representative data based on correlating sampled data spatially is a common practice. In this paper a correlation method is introduced, the correlation between a sensor node's data and its neighboring sensor nodes' data is measured. Based on this method a data correlation clustering method is presented, where a representative node is elected and sends its observation to the sink, thus reducing the number of transmissions to the sink.

Keywords: *Wireless sensor network, Data aggregation, Clustering method, Data density*

1. Introduction

A WSN typically has little or no infrastructure. It consists of a number of sensor nodes (few tens to thousands) working together to monitor a region to obtain data about the environment. There are two types of WSNs: structured and unstructured. An unstructured WSN is one that contains a dense collection of sensor nodes. Sensor nodes may be deployed in an ad hoc manner into the field. Once deployed, the network is left unattended to perform monitoring and reporting functions. In an unstructured WSN, network maintenance such as managing connectivity and detecting failures is difficult since there are so many nodes (J. Yick *et al*, 2008).The main goal of a WSN is reliably detecting and accurately evaluating the proceedings in the monitored area with collected data.

A WSN composed of self-organized wireless sensor nodes scattered in a monitored area collects, processes and transmits data acquired from the physical atmosphere (J. Yick *et al*, 2008; L.M. Oliveira *et al*, 2011). The main goal of a WSN is constantly detecting a precisely evaluating the events in the monitored area with the collected data. For this purpose, sensor nodes should be deployed closely. However, this will cause overlapping of sensor nodes' sensing areas and the spatial idleness of adjacent sensor nodes' data (J.N. Al-Karaki *et al*, 2009). If every sensor node transmits collected data to the sink node, the sensor nodes will use a large quantity of energy. To reduce the amount of transmitted data in a WSN, a great number of correlation based data aggregation methods have been studied in the literature (G.A.Shah *et al*, 2007; C.Hua *et al*, 2008).

In the study of data aggregation methods, the spatial correlation model between sensor nodes' data is an main establishment that relates to the precision of aggregated data and energy utilization of sensor nodes (J. Yuan *et al*, 2009).The energy effectiveness of the DCCM clustering method is used in data transmitting process. While in the clustering process, the DCCM clustering method is an energy efficient one. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor networks (WSN) offer a progressively more attractive method of data gathering in distributed system architectures and dynamic access via WSN.

2. Related Work

Coverage problem in WSN

(G. Fan *et al*, 2010) The most important performance metrics for sensor networks shows the efficient way of

*Corresponding author **Savita Patil** is a M.Tech Scholar and **Sunita Rawat** is working as Assistant Professor

monitoring a sensor field. Each sensor coverage model is dependent on their sensing function, while the coverage over an entire network is a collaborative performance metric for the sensor nodes deployed over a geographical region. The main objective of article survey and research progress is made to give an overview of coverage problems faced in wireless sensor networks. The sensor networks coverage problems can be classified into three different types based on the topic to be covered (J. Yuan *et al*, 2009). The article also presents the review representative solution approaches. Hence overhead is high and power consumption is more.

Distributed clustering-based aggregation algorithm

(Yajie Ma *et al*, 2011).proposed a α -local spatial clustering algorithm for sensor networks. By evaluating the spatial connection between data gathered by multiple sensors, the algorithm forms a commanding set as the sensor network strength, used to recognize the data aggregation.

This algorithm is useful for applications such as environmental surveillance where the sensors are always distributed in very high density. Discussed time and message complexities of algorithm with analysis of size of aggregated networks. The trial consequences show that the aggregated network can provide the environmental information in very high accuracy in comparison with the original network. Result of average percentage membership match (APMM) of α-LS clustering is high other algorithms are less which is much lower than APMM of this algorithm. Therefore algorithm has better accuracy performance in data description/summarization (R.K Shakya *et al*, 2012)

Slepian-wolf coding in cluster-based WSN

(J. Zheng *et al*, 2010) studied the major problems in applying Slepian-Wolf code for aggregating drain cluster-based wireless sensor networks. First judge the clustered Slepian-Wolf coding (CSWC) problem, which aims at choosing a set of displace possible clusters to cover up the entire network such that the general density gain of proposed code is maximized, and also proposes a distributed optimal-compression clustering (DOC) protocol to solve the problem. Under a cluster chains of command developed by the DOC protocol, and then think the optimal intra-cluster rate-allocation problem. Prove that there exists and optimization algorithm that can find an best rate allocation within each cluster to reduce the intra-cluster communication cost and present an intra-cluster coding protocol to locally execute Slepian-Wolf coding within a only cluster but it has inadequate network connectivity and inadequate number of clusters can be produced.

Energy efficient routing protocol based on MAC protocol

(Rajeswari *et al*, 2011). For wireless sensor networks using spatial correlation medium access an Energy efficient routing protocol is proposed and studied in comparison with IEEE 802.11. In Wireless Sensor Networks, an energy efficient medium access control protocol is required for obtaining higher energy efficiency in very difficult operating conditions, where node and link failures are common. The MAC protocol manages radio transmissions and receptions on a shared wireless medium. Node selection procedure is used to recognize the core sensor node ensuing in least amount of energy utilization in WSN. For that, Iterative Node Selection (INS) algorithm is introduced in this paper. Considerable development is achieved in distortion by selecting the location of core sensor node using vector quantization (VQ) algorithm.

3. Clustering Method Description

Data Correlation Clustering Method

In a WSN, if a certain number of neighboring sensor nodes' data are close to a sensor node's data, this sensor node can represent its neighbours in the data domain. This representative sensor node is called the core sensor node.

Core sensor node: Let us consider sensor node v has n neighbouring sensor nodes. They are respectively v 1 , v 2 . . . v n . The data object of v is D. It's neighbouring sensor nodes' data objects are respectively D1 , D2 . . . Dn . If there are N data objects in D1, D2 , . . ., Dn whose distances to D are less than ε then the sensor node v is called the core sensor node. Where ε is the data threshold.

Instinctively, the larger the N is, the improved representative the sensor node v is to those sensor nodes whose data objects are in ε-neighbourhood of D. Meanwhile, high interest of the data objects in the ε-neighbourhood of D implies that sensor node v has a high spatial correlation between it and these sensor nodes. Therefore, to determine the representation degree of v to those sensor nodes whose data objects are in ε-neighbourhood of D in quantity, we proposed the data correlation clustering method as shown in below.

Data correlation clustering method

Let sensor node v has n neighboring sensor nodes which are inside the cycle of the communication radius of v. They are v 1 , v 2 . . .v n , respectively. The data object of v is D, and its neighbouring sensor nodes' data respectively D1, D2 . . . Dn. Among these n data objects, there are N data objects whose distance to D is not as much of than ε. After that the data correlation clustering method of sensor node v to the sensor node whose data objects are in ε-neighbourhood of D is as follows.

$$\mathrm{Sim}(V_1)=a_1\left(1-\frac{1}{\exp(\mathrm{N})}\right)+a_2\left(1-\frac{\mathrm{d}\Delta}{\epsilon}\right)+a_3\left(1-\frac{\mathrm{d}}{\epsilon}\right) \quad (1)$$

Where ε is data threshold. dΔ is the distance between D and the data centre of the data objects which are in the ε neighbourhood of D. d is the average distance between the N data objects and D. a1+a2+a3=1.

If the data correlation clustering method of sensor node v is Sim(V_1) defined by Eq.1, then we can obtain the properties of Sim(V_1) as.

1) Sim(V_1) increases with the increase of N, the amount of data objects which are in the ε-neighbourhood of D;
2) Sim(V_1) increases with the increases with decreases of dΔ, the distance between D and the data centre of the data objects which are in the ε-neighbourhood of D;
3) Sim(V_1) increases with the decreases of d, the typical distance between D and those data objects which are in the ε-neighbourhood of D; 4) Sim(V_1) ϵ[0,1].

These properties are consistent with our intuitiveness. In data correlation clustering method, the data threshold ε guarantees that Sim(V_1) will not be impacted by unrelated data. The amount threshold min Pts is the minimum amount for sensor node v to represent some sensor nodes. In order to exhibit the potency of data density degree defined by Eq.1. It includes three procedures: the Sensor type calculation (STC), Local cluster construction (LCC), and Global representative sensor node selection (GRS).

4. Simulation Results

The simulation studies involve the network topology where we consist of 49 nodes in the network which as shown in the Fig 1. We are performing simulation using NS-2. After the nodes are deployed in the network local clusters are formed by judging core sensor node as shown in Fig 2.Identifying member of sensor nodes by sending information to its neighbours as shown in Fig 3.Core sensor node is elected as shown in Fig 4.merging local clusters to form global cluster as shown in Fig 5.Hence we calculate the energy consumption as shown in Fig 6.packet delivery ratio as shown in Fig 7 using this simulation and finally obtain graphs. We calculate how much energy is consumed and compare it with the other algorithm as well as the proposed one then the graph is obtained.

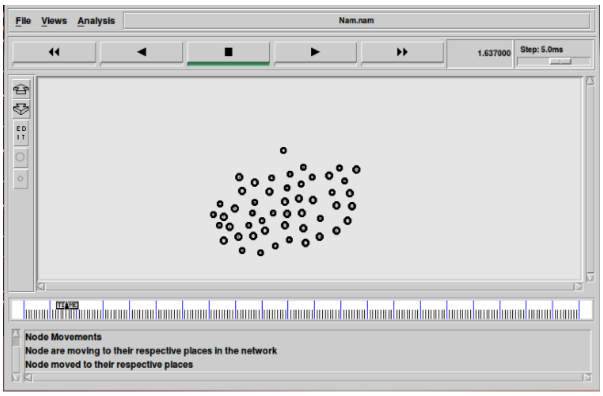

Figure 1: Deployment of nodes

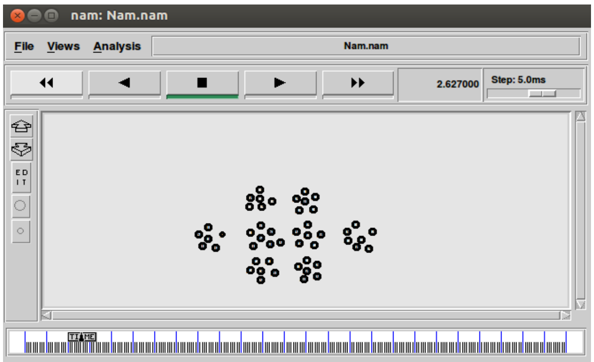

Figure 2: Forming local cluster

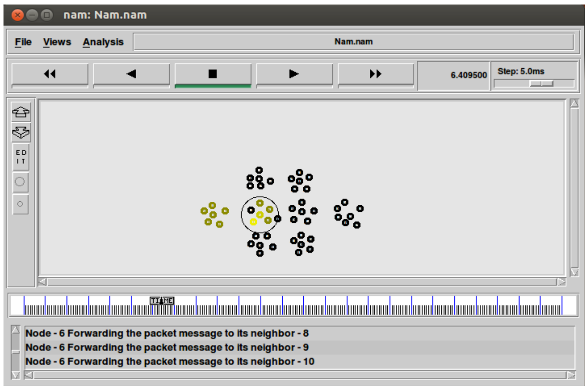

Figure 3: Identification of members of local cluster

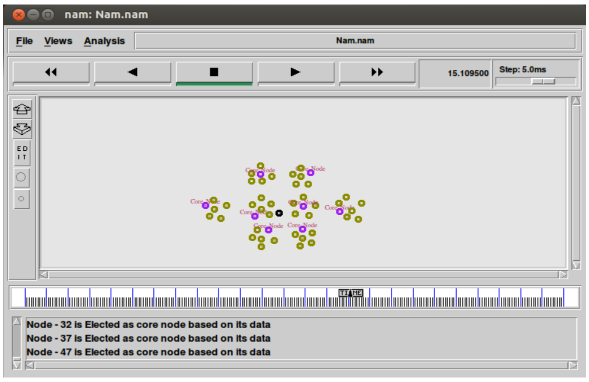

Figure 4: Selection of core sensor node

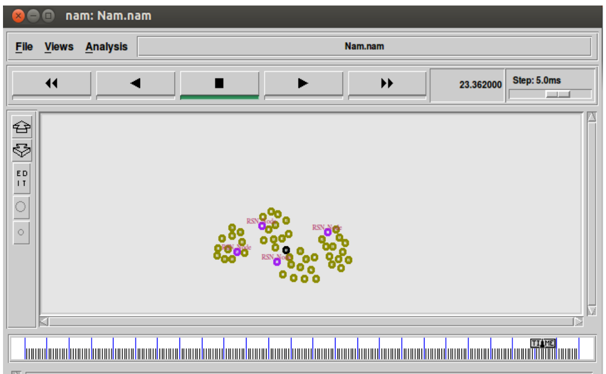

Figure 5: Merging local clusters

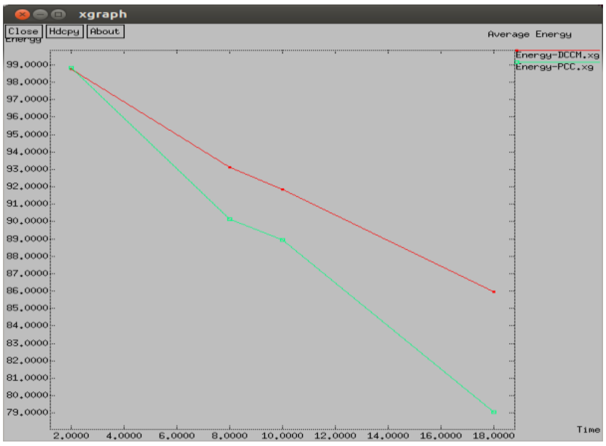

Figure 6: Energy consumption

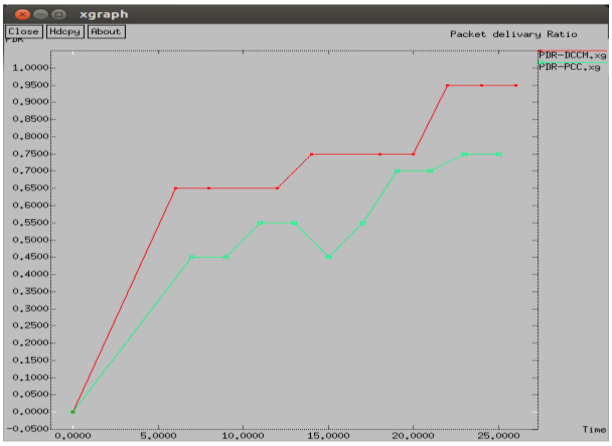

Figure7: Packet delivery ratio

Conclusion and Future Work

In this paper, the data correlation method focuses on correlating data in complex environment. The Data correlation clustering method is introduced, with this method the sensor nodes that have high correlation are in the same cluster, a representative sensor node is elected from merged cluster, and this representative sensor node sends data to the sink node.

Because of using high correlation of sensor nodes, accurate data can be achieved and by merging small clusters the numbers of transmissions are reduced. The direct further work on the data correlation clustering method is, the performance of the data aggregation protocol is coupled with the infrastructure of the network and it is also required to analyse the possibility of including the security protocols to prevent the data from attacks. It would be interesting to extend the work to more general network topologies.

References

J. Yick, B. Mukherjee (2008), and D. Ghosal, Wireless sensor network survey, Comput. Netw, 8 (12), pp 2292–2330.

L. M. Oliveira and J. J. Rodrigues (2011), Wireless sensor networks: A survey on environmental monitoring, J. Commun, 6 (2), pp 1796–2021.

J. Yuan and H. Chen (2009), The optimized clustering technique based on spatial-correlation in wireless sensor networks, Proc. IEEE Youth Conf. Inf, 4 (7), pp 411–414.

G. A. Shah and M. Bozyigit (2007), Exploiting energy-aware spatial correlation in wireless sensor networks, in Proc. 2nd Int. Conf. Commun. Syst. Softw. MiddleWare, 5 (2), pp 1–6.

C. Hua and T.-S. Yum (2008), Optimal routing and data aggregation for maximizing lifetime of wireless sensor networks, IEEE/ACM Trans. Netw, 16 (4), pp 892–903.

J. N. Al-Karaki, R. Ul-Mustafa, and A. E. Kamal (2009), Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms, Comput. Netw, 53 (7), pp 945–960.

G. Fan and S. Jin (2010), Coverage problem in wireless sensor network: A survey, J. Netw, 5 (9), pp 1033–1040.

Y. Ma, Y. Guo, X. Tian, and M. Ghanem (2011), Distributed clustering-based aggregation algorithm for spatial correlated sensor networks,IEEE Sensors J, 11 (3), pp 641–648.

J. Zheng, P. Wang, and C. Li (2010), Distributed data aggregation using Slepian-Wolf coding in cluster-based wireless sensor networks, IEEE Trans. Veh. Techno, 59 (5), pp 2564–2574.

A. Rajeswari and P. Kalaivaani (2011), Energy efficient routing protocol for wireless sensor networks using spatial correlation based medium access control protocol compared with IEEE 802.11,in Proc. Int. Conf. PACC, 8 (2) pp 1–6

R K. Shakya, Y. N. Singh, and N. K. Verma (2012), A novel spatial correlation model for wireless sensor network applications, in Proc. 9th Int. Conf. WOCN, 3 (7) pp 1–6.