Management of Big Data for Improving Performance of Wireless Sensor Network Using a Structured Clustering Approach

Gauri Kalnoor and Jayashree Agarkhed

Abstract--- *Monitoring variety of temperature and ambient humidity is one of the most important and widespread application of Wireless Sensor Networks(WSN). Thus, the supply of energy for each sensor node in WSN is very limited which results in the reduction of energy usage in many applications, increasing the lifetime of WSN. However, a well-organized sensing data that is collected in large amount and distributed densly over the wide network of sensors is the major challenge in most of the research areas. The mobility of a sink node is considered to be an effective solution for managing big data and its collection. Here, we develop a framework to leverage the correlation between set of sensor nodes that are most active.*

Keywords--- *Big Data, WSN, Mobile Sink, Routing, Energy Consumption, Spatial Correlation.*

I. INTRODUCTION

APPLICATIONS that are most widespread and well-suited in WSN include smart spaces, environmental monitoring, battlefield surveillance, biological detection and so on. Based on the way data is collected, the applications are classified into two broad categories known as continuous sampling data and event based detection. The data is expected to be collected by the sensors and report to the sink only when an event of interest is occurred [1], thus leading to error sensitivity and intolerance delay in WSN. This characteristic refers to event based detection of data collection. In continuous sampling data, the sensors regularly samples and reports the data to the sink where the snapshot of area for the collected data is generated [2].

In WSNs, the applications should be designed such a way that the problem of redundant data and energy constraint should be overcome, especially when considering high density area of network [3]. Reduction in amount of communication with maintaining integrity is one of the most challenging issue of WSN, achieving longer lifetime. Wireless Communication results in significant amount of consuming power during delivery of data, due to very low consumption of energy for sensing and computation. The data set is analysed that infers the node redundancy with its degree. The best approach is to schedule the sensor nodes in sleep mode that results to save energy.

Some examples for big data include data of mobile phone, twitter data, queries online and so on. The digital data that is massively generated in large amount with high frequency [4] is referred with the term "Big Data". Some features of Big data are as detailed below:

1) Generation of digital data: Data is generated digitally and stores it in the form of one's and zeroes series. This can be altered by computers.
2) Collection of data automatically: Data that is redundant whenever generated is extracted and stored in a place in system.
3) Trackable geographically or temporally: data location is trackable with time needed for generating data.
4) Produced passively: Data produced passively is digital services to be interacted.
5) Continous analysis: Information relevant can be developed and analysed in real time.

The sensors in WSN are developed such that they relay the data gathered to the sink node. The section below describes about the work done by many researchers in the field of big data.

II. RELATED WORK

The large amount of data collection process becomes easier for acquiring energy effiecient network uses a mobile sink approach. The authors explains in [5] the data gathering approach using a sink which is mobile, and the network is divided into number of clusters. The centeroid node of each cluster forms the sink node to manage big data of the sensor network.

In [6], the authors have proposed an algorithm called as Expectation Maximization(EM), where centroid of the cluster is selected based on the distance of wireless communication. It addressess the mobility of each sensor node, with the changes in distance parameter in WSN.

The authors in [7] have proposed a new algorithm for dynamic network of sensors where data can be gathered in an efficient manner for energy consumption. Analysis of big data is the core of modern business and science. The three important charecterstics of big data discussed by the authors in [8] are velocity, variety and volume. The study of data streams that are generated by the sensor nodes to embed in the machinery for improving performance of the products. The

Gauri Kalnoor, Assistant Professor, Computer Science and Engineering, Central University of Karnataka, Kalaburagi, India. E-mail: kalnoor.gauri@gmail.com

Jayashree Agarkhed, Professor, Computer Science and Engineering, PDA College of Engineering, Kalaburagi, Karnataka. E-mail: pdajayashree@gmail.com

DOI:10.9756/BIJSESC.8232

authors explains that the large amount of data generated is monitored consistently to achieve safety, quality, maintainance, security and so on.

In [9], the authors have discussed the method to monitor the data collected in healthcare services which is considered to be a rich source of big data. The authors in [10] presented a framework for services of sensors mainly a cloud based network. WSNs may be divided into sub- networks because of many limitations in wireless communications. One of the most emerging and prominenet feature is the scheme using mobile sink.

In [11], the presenters have discussed one of the popular algorithms called LEACH (Low-Energy Adaptive Clustering hierarchy). This algorithm includes static sink and is executed by each sensor node in LEACH. The information exchange takes place by each sensor nide and the node with high residual energy has the probability much higher compared to other nodes for getting elected as a cluster head. Energy consumed by each node is maintained equally, due to periodic re-clustering as discussed in [12].

An algorithm known as k-CONID was proposed by authors in [13] and considered to be a probabilistic algorithm. Each node exchanges their respective random IDs and the one that acquires minimum ID within the range of k-hop becomes the cluster head of the cluster. This algorithm minimizes the transmission of data thus achieves optimization and minimum energy consumption.

The k-means algorithm explained in [14] is the centralized clustering algorithm where a super node conducts the scheme of mobile sink. The mobile sink may fail to gather information from all the sensor nodes, and thus an algorithm called as KAT mobility (k-means and TSP mobility) was proposed in [15].

III. NETWORK MODEL FOR BIG DATA MANAGEMENT

The network with number of sensor nodes would require management of big data at each level. The big data for each node comprises of high velocity, high volume and information asset with high variety [16]. This type of data is difficult to collect, store and then process the same to the destination node based on the available technologies to be used. The data that is generated with sources of wide range such as Radio-Frequency-Identification, Machine-to-Machine and so on are structured with high variety of information. The high volume refers to large amount of data that needs to be analysed and processed in the specified duration of time.

In distributed WSN, the dense collection of data requires to achieve energy efficiency. For example, the data transmitted is capable of shrinking its desired volume using the technology of data compression [17]. This technology needs each of the nodes equipped with storage capacity of large volume and thus increasing the power computational to much higher level. Redundant wireless transmission is reduced by using best topology technique. Figure 1 [18] shows the major trends of gathering the big data to improve performance of WSN.

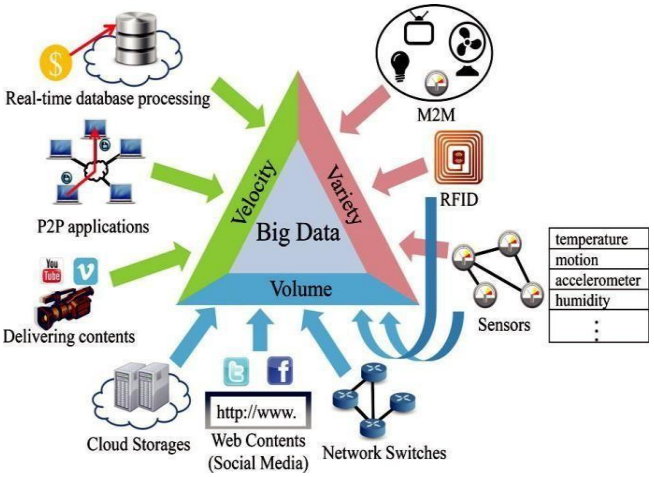

Figure 1: Trends of Big Data Collection

A. *Design of a Network Mode*

The network in our model to be considered is comprised of a sink node that is mobile in nature and large number of sensor nodes are spread within the limited range of network. The network is divided into number of sub- networks called as clusters, where the nieghbouring nodes within an expected range is considered. A localization technology is used by every sensor nodes to know its own locations. The mobile sink also knows the locations of each other sensor nodes by using this technology. The sensors as well as a sink has limited range of communication called as R and a successful communication is considerd when it is within the range R. The centroid node of each cluster after clustering formation is patrolled by the sink which is calculated to obtain minimum consumption of energy during transmission of data. A buffered memory is equipped within each sensor node to store the sensed information until the sink reaches the centroid of the cluster. K centers of clusters visit the mobile sink with information exchange between nearest neigh bouring nodes in the target area of L × L.

Figure 2 shows the network model designed for after cluster formation takes place.

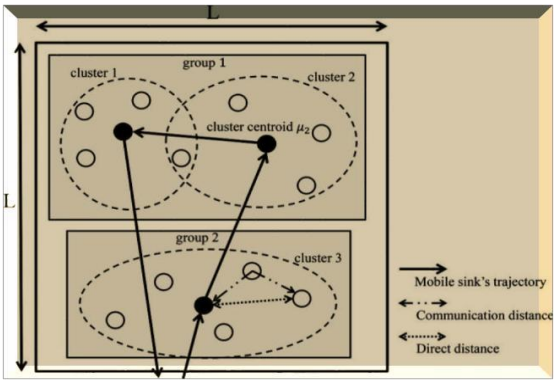

Figure 2: Network Model for Big Data

In the figure above, the clustering formation for the scheme of mobile sink needs to choose the best algorithm used to divide the nodes into number of clusters and number of optimal clusters to reduce energy consumption.

B. *Algorithm of Big Data in Mobile Sink*

The algorithm is designed as a modified clustering algorithm for gathering big data using the scheme of mobile sink.

Algorithm Mobile Sink Scheme

Step 1 : Start
Step 2: Node deployments using heterogeneity of WSN
Step 3: do node broadcasting to all nodes within its range
Step 4: while(high power node N= receive broadcast _)
Reply sent to the node of the cluster$_i$
End while
Step 5: All the replies from each node are compared from the high power node and the b cluster head is chosen among the set of high power nodes.
Step 6: Node Selection broadcast
Step7: Cluster formation takes place with node notification.
Step 8: Static sink is sekected in each cluster to gather huge amount of data from each node.
Step 9 : Cluster head counts the number of packets transmitted between each node and informs the mobile sink, when threshold value is reached.
Step 10: The data is sent to the base station
Step 11: While (nodes in cluster $_i$ == NULL)
Cluster head changes to sleep mode
Mobile sink gathers information of high volume
End While
Step 12: End

In this algorithm, the cluster formation takes place with election of a cluster head. At this stage, the cluster nodes elects a sink node which is constantly moving such that information of large volume can be gathered from each node and then the level of packets is checked such that it does not reach the threshold. Once any node leaves the cluster , the information is sent from the cluster head to the sink node and then to the base station. The cluster is deformed when all the nodes leaves the cluster network.

IV. PERFORMANCE RESULTS

In this section, we discuss the performance evaluation of the network using the proposed approach. Figure 3 shows the framework of network model [19] to improve performance measures.

Figure 3: Framework of Network Design

A. *Collection of Big Data*

The big data gathering follows with the formation of clusters and sensors those are active are selected as samples. The fixed number of sensor nodes are determined in the entire phase of data collection. The parameter is declared globally denoted by , the time period for collecting data. The parameter is the time period of learning the gathered data by the mobile sink.

The energy consumed by sampler nodes which are active is more compared to that consumed by the nodes in sleep mode.

A feasible solution for adjusting the time period for data collection is that the 'relearning" message needs to be broadcasted over all the nodes in WSN by the cluster head nodes.

B. *Energy Consumption*

The consumption of energy is analysed based on the number of sensor nodes that are sleeping/working iin every cluster in data collection phases and learning phases. The readings is sent to the cluster head nodes by all the member nodes in the learning phase, and no node is in sleeping mode.

The cost of energy with nodes N in a cluster during the learning phase is as given in equation 1.

$$E_l = {}_lN \text{----------(1)}$$

Where e is the cost of energy when data transmission between pair of nodes takes place.

In data collection phase, the cost of energy in every cluster is determined using equation 2.

$$E = TN_a \text{ ----------(2)}$$

Where N_a is the number of active nodes.

The energy consumed for the unscheduled nodes is compared and the percentage of energy saved during the the complete cycle of learning and data gathering phase, within the cluster is estimated using equation 3.

$$P_E = - \frac{T_l N + T N_a}{(T_l + T\ )N} \text{ ----------(3)}$$

The energy comsumption is reduced when the period of data collection T is extended. Thus, the number of nodes that are active need to be reduced to achieve performance of the network.

The data analysed at sink node to achieve better performance is evaluated using equation 4.

$$ERq = KRC \text{----------} (4)$$

Where E is the energy consumed while gathering data

K is the number of clusters

N is the the number nodes in each cluster with high volume of data

R is the data needed to be stored and processed

C is the data transmitted

The parameters are set using equation 3 as follows:

N= 500, t=0.4 , T_l= 50, T=4T_l=200
Active Nodes : 35%
P_E= 51.04%

The results for the scheme to be proposed is tabulated

based on the experiments conducted in the previous research work by different authors as shown in table 1.

Table 1: Results

Number of Clusters	*Amount of Data stored/ Cluster*	*Energy Consumed*
10	5 GB	40%
15-50	20 GB	60%

V. CONCLUSION

The efficiency of energy is examined where data is gathered method by using sink mobility approach in WSN. The methods used in the existing system in WSN are not applicable for dynamic sensor nodes. Hence for data collection from the dynamic WSN, the sink approach was proposed. Moreover, the clustering algorithm in modified version is applied for the big data storage and improve the performance of the network. This mobile sink approach can also applied to homogeneous dynamic wireless sensor network which is considered as an open problem.

REFERENCES

[1] D.Takaishi, H. Nishiyama, N. Kato and R. Miura, "Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks", IEEE Transactions on Emerging Topics in Cloud Computing, 2014.

[2] Yi sun, Can Chi, Shanshan Ke and R. Jun Lu, "Esearch on Dynamic Clustering Routing Considering Node Load for Wireless Sensor Networks", Communications and Network, Vol. 5, Pp. 508-511, 2013.

[3] P. Krishnan, "Comparison and Performance Analysis of Dynamic and Static Clustering Based Routing Scheme in Wireless Sensor Network", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, No. 4, 2013.

[4] K.R. Rashmi, A.B. Shivakumar and J. Ananda Babu, "Energy Efficient Clustering Techniques For Mobile Data Gathering In Distributed WSN", International Journal of Advanced Technology in Engineering and Science Vol. 3, No. 1, 2015.

[5] P. Sasikumar and S. Khara, "K-Means Clustering In Wireless Sensor Networks", International Conference on Computational Intelligence and Communication Networks, 2012.

[6] A.B.M. Alim Al Islam, C. Sayeed Hyder H. Kabir and M. Naznin, "Stable Sensor Network (SSN): A Dynamic Clustering Technique for Maximizing Stability in Wireless Sensor Networks", Wireless Sensor Network, Vol. 2, Pp. 538-554, 2010.

[7] K.R. Remesh Babu and G.J. Suja, P. Samuel and S. Jose "Performance Analysis Of Big Data Gathering in Wireless Sensor Network Using An EM Based Clustering Scheme", International Conference on Advances in Computing and Communications, 2015.

[8] D. Virmani, A. Jain, A. Khandelwal, D. Gupta and N. Garg, "Dynamic Clustering Protocol for Data Forwarding in Wireless Sensor Networks", International Journal of Computers and Technology, 2013.

[9] S. Benkirane, A. Beni hssane, M. Lahcen Hasnaoui, M. Saadi and M. Laghdir, "Weighted Dynamic Distributed Clustering Protocol For Heterogeneous Wireless Sensor Network", International Journal of Wireless and Mobile Networks(IJWMN) Vol. 4, No. 6, 2012.

[10] U. Raza, A. Camerra, A.L. Murphy, T. Palpanas and G.P. Picco, "What does model-driven data acquisition really achieve in wireless sensor networks? In: IEEE International Conference on Pervasive Computing and Communications, 2012.

[11] H. Jiang, S. Jin and C Wang, "Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks", 2011.

[12] W. Wu, H.B. Lim and K.L. Tan, "Query-driven data collection and data forwarding in intermittently connected mobile sensor networks", 2010.

[13] D. Brunet, E.R. Vrscay, Z. Wang, "On the mathematical properties of the structural similarity index", Vol. 21, Pp.1488–1499, 2012.

[14] C. Zhu, C. Zheng, L. Shu and G. Han, "A survey on coverage and connectivity issues in wireless sensor networks", Vol. 35, Pp. 619–632, 2012.

[15] L. Ramaswamy, V. Lawson and S. Gogineni, "Towards a quality-centric big Data architecture for federated sensor services", In IEEE International Congress on Big Data , 2013.

[16] S. He, J. Chen, D. Yau and Y. Sun, "Cross-Layer optimization of corre-lated data gathering in wireless sensor networks", In IEEE Communica-tions Society Conference on Sensor Mesh and Ad Hoc Communications and Networks, 2010.

[17] I. Bisio, M. Cello and M. Davoil, "A survey of architectures and scenarios in satellite-based wireless sensor networks", System de-sign aspects, International Journal of Satellite Communications and Networking, Vol. 30, No. 6, 2012.

[18] Oracle, "ÒBig data: Business opportunities", requirements and oracles approach, Pp. 1Ð8, 2011

[19] S. Sagiroglu and D. Sinanc, "ÒBig data: A review", In International Conference on Collaboration Technologies and Systems (CTS), 2013.

Gauri Kalnoor, received her B.E, Mtech in Computer Science and Engineering from the reputed university. Currently pursuing PhD in the area Wreless Sensor Network. (E-mail: kalnoor.gauri@gmail.com)