VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

CLUSTER BASED PRIORITY TRAVERSE PROCESSING FOR EFFICIENT DATA COLLECTION IN WSN

S. Padma Priya and E. Srie Vidhya Janani
Department of Computer Science and Engineering, Anna University regional center, Madurai, India
E-Mail: priyaveeracs@gmail.com

ABSTRACT

Wireless Sensor Networks (WSNs) embodies sensor nodes, which are sufficiently little to deploy where legitimate base is not accessible. With an imperative of constrained battery power, information accumulation has a noteworthy effect in vitality utilization of sensor nodes. In our proposed strategy Cluster-Based Priority Traverse method, portable sink cross in an element way bound by the static sink, which chooses the way from the prioritization routing table. Information that are sensed are not all of much significant. Sensed information is arranged into essential information, information that shows radical variety from past information and auxiliary information, information that fit in with typical sensed information pattern. clusters in the system are organized based upon their Primary Data Count. Classification of information is performed by computing mean and standard deviation for the sensed within a group and discovering the information of extensive variety. To give high dependable navigate way to mobile sink to accumulate information from groups with high need sooner than low need cluster. Minimum Spanning Tree is developed for each group to all other clusters in the system. The performance of the network is investigated through various simulations.

Keywords: wireless sensor networks (WSN), cluster-based priority traverse (CBPT), primary data count (PDC).

1. INTRODUCTION

Wireless Sensor Network has an adventurous enhancement in disaster recovery applications. It plays an important role in forest fire detection as in [9]. There exist many researches to make fire detection more reliable. Temperature and humidity that are sensed by sensors are not normal when there is even a small traces of smoke. At that instance, the mobile sink that traverses a static path around the network is able to gather the data from the smoke originated region only when the mobile sink reaches the transmission range of that particular node. Treating all the data of equal preference may lead to delivery of event-driven data with latency. Interesting events are to be notified by the sink without latencies. Primary data detection in wireless sensor networks (WSNs) are those data instances that deviate from the rest of the data patterns based on a certain measure. Secondary data are the belongings of same phenomenon are spatially correlated. The nodes near a sink can be burdened with relaying a large amount of traffic from other nodes. This phenomenon is sometimes called the “crowded center effect” or the “Hot Spot problem”. Mobile sink overcomes the crowded center effect that caused by static sink. Mobile sink is a moving element, where the robot or moving vehicle acting as the mobile base station, moves on a predetermined path. Mobile sink has advantages of distributing the load, collecting data continuously and moves slowly and discontinuously as in [5]. CBPT is a novel traversing technique in which mobile sink node follows the data collection path decided by the static sink, which is an efficient technique for the data collection from the clusters. The data collection using prioritized table is suitable routing for delay sensitive data collection network structure.

This technique initiates by dividing the network into optimum number of clusters with optimum size as in [2]. Distributed implementation of centralized algorithm is slightly revised by rotating the cluster head role among the cluster head election candidates with high residual energy and backbone tree is constructed with new selected CH. Within this meantime the energy drained out CH regains energy through energy harvesting. This revised version of the algorithm reduces the message overhead of electing new CH at an instance of unavailability of existing cluster head. Each cluster head identifies those measurements that significantly deviate from the normal pattern of sensed data by computing the median of its member’s sensed data then it finds the difference between individual sensed data and median. Finally cluster sample mean and sample standard deviation from the differences. From the final standardized value the primary data is detected. A sensor is data of interest if the absolute value of its standardized difference is sufficiently large.

Related work is presented in Section II. Section III provides cluster formation. We show revised distributed implementation of centralized algorithm in Section IV. Categorization of primary and secondary data is presented in Section V followed by Section VI that provides CBPT algorithm based on prioritization routing table. Results are discussed in Section VII and we conclude in Section VIII outlining some directions for future work.

2. RELATED WORKS

Several works have considered the opportunistic utilization of mobile sink. Many priority-based algorithm have considered buffer overflow as in [3], time delay as in [1] as their metric to prioritize the sensed data. The concepts they left to point out is, considering buffer size alone as a prioritizing metric is unaware of the data importance. Early deadline first (EDF) considers time as a routing preference metric, which can route the ordinary spatially correlated data that approaches the time deadline earlier than the event-driven data. This algorithm is suitable for clock-driven application than event-driven application. Preferred data collection in a network with mobile sink following a static routing path method will experience a delay as in [4], this algorithm is only suitable for delay tolerant applications. A dynamic sleep time control in event-driven network, which does not provide different priorities traffic services as in [8]. Priority based hybrid routing as in [6] proposes a new approach of geographic diffusion but with a limitation of static sink.

When a remarkable change in the readings of sensors is detected, an outlier or some event must have occurred. This observation is explored as in [10], [12], [11] for 0/1 decision logical computation. The related algorithms require only the most recent sensed data of individual sensors. No collaboration among neighboring sensors is exploited. As in [10], the change points of the time series are statistically computed. The detector proposed as in [12] computes a running average and compares it with a threshold, which can be adjusted by a false alarm rate.

3. CLUSTER FORMATION

In data gathering clustering has much advantage than routing tree structure. Clustering has better traffic load and efficient energy consumption. The network is divided into optimum number of optimum sized cluster calculated as in [2]. The data flow of cluster selection algorithm is shown in Figure-1.

Nodes within Hr range is selected as election candidate. The value of H is determined such that there is at least one node within H hops from the centroid of a cluster and r is the transmission range of the sensor.

4. REVISED CENTRALIZED ALGORITHM

As in [2] the distributed implementation of centralized algorithm states that the if the CH drains out of energy new CH is selected by initiating the same CH selection algorithm, which in turn increase the message complexity of the algorithm.

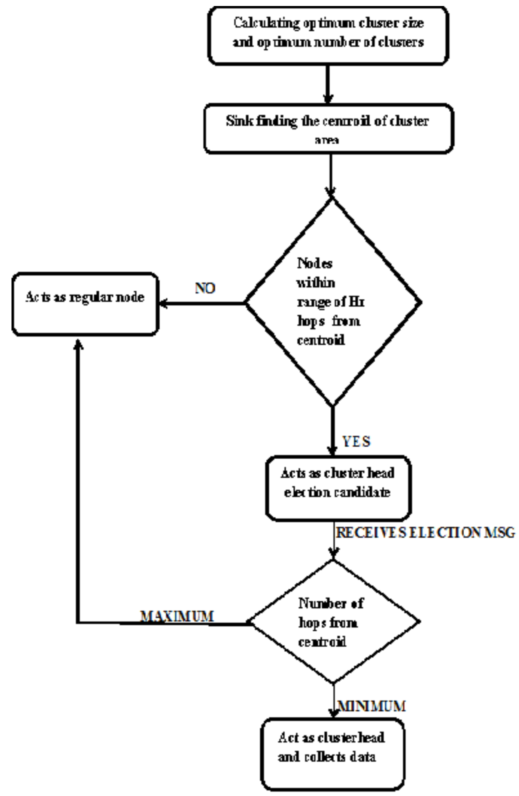

Figure-1. Data flow of CH selection algorithm.

Revised algorithm claims that the any election candidate within the Hr subset with highest residual energy will be the preceding CH of the cluster. A new backbone tree is constructed by including the new CH. Let EC={C_1, C_2…..C_i} be the subset of election candidates. C_j be the next nearest node to the centroid of the cluster with high residual energy then it becomes the CH for the remaining round of transmission when CH_c runs out of energy.

Revised algorithm
1. If e (CH_c)=0
2. And Let EC=e {C_1,C_2,..Ci}!=0
3. Then Cj ϵ EC, next nearest node to centroid
4. With e (C_j) > {e(EC)}
5. CH_c=EC_j; else Repeat 3 and 4 until j=i
6. End

From this algorithm j ranges from 1 to i. e(C_i) represents the energy level of the election candidate. CH_c is the cluster head for current round of transmission.

5. CATEGORIZATION OF DATA

The point of separation of the outlier data and the event-driven data is important here. "An outlier or anomaly is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data"[7]. Actual sources of outliers in data collected by WSNs include noise and errors, actual events. Primary data detection also provides an efficient way to search for values that do not follow the normal pattern of sensor data

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

in the network. The detected values consequently are treated as events indicating change of phenomenon that are of interest. On the other hand, the common characteristic of outlier detection and event detection is that they employ spatiotemporal correlations among sensor data of neighboring nodes to distinguish between events and noises. This is based on the fact that noisy measurements and sensor faults are defined to be unrelated, while event measurements are likely to be spatially correlated. Since sensor nodes are uniformly and independently deployed in a rectangular sensor field within a transmission range of sensor nodes is r. That is, any two sensors whose Euclidean distance is within r can communicate with each other. So the sensor nodes are spatially correlated and there is no chance of occurrence of erroneous data in the network. The data with drastic change with predefined past data are confidently event-driven data.

Consider there are m is the optimum number of clusters with n number of sensor nodes. Let C_i be the subset of clusters where i={1,2,…m} and S_{ij} be the subset of sensor within a cluster i where j={1,2,…n}. CH of the cluster C_i gather the sensed measurements $\{SM_1, SM_2,\dots SM_n\}$ from its cluster members. CH arranges the measurements in increasing order. Let med_i be the median calculated from the measurements of cluster i by using Equation (1).

$$med_i = \begin{cases} \left(\frac{n+1}{2}\right) th\ observation, & odd\ n \\ \frac{n}{2} th + \left(\frac{n}{2}+1\right) th\ observation, & even\ n \end{cases} \quad (1)$$

Median more accurate than mean because the drastic change the dataset will deviate the mean. This is because the sample mean cannot represent the center of a sample well when some values of the sample are out of predefined range. However, median is a accurate estimator of the center of a sample.

Let D_j be difference between the SM_j and med_i by using Equation (2) That is the difference between the sensed measurement of j^{th} sensor node present in the cluster i and the median of cluster i.

$$D_j = SM_j - med_i \quad (2)$$

Let μ_i be the average of the difference D_j and σ_i be the standard deviation of the sample which is calculated by Equation (3) and Equation (4)

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n}(D_j) \quad (3)$$

$$\sigma_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(D_j-\mu_i)^2} \quad (4)$$

Where n is the total of sensor nodes in the cluster. Standardize the data set to obtain $\{X_1,X_2,\dots X_j,\dots X_n$ by Equation (5).

$$X_1=((D_1-\mu_i)/\sigma_i),\dots X_j=((D_j-\mu_i)/\sigma_i),\dots X_n=((D_n-\mu_i)/\sigma_i) \quad (5)$$

If $|X_i| \geq$ TL, then the measurement of i^{th} sensor is considered as primary data and the remaining data are secondary data.TL is threshold limit which is obtained from the test data is compared against the learned predictive model for normal or abnormal classes. This predefined threshold is also obtained from previous historical pattern.

6. CBPT ALGORITHM

CBPT algorithm initiates with the primary data count (PDC) from the cluster heads of the network. CHs maintain PDC is initialized to be 0. It gets incremented whenever the CH encounters a primary data based on the classification method discussed in section 4. Since a backbone tree exists through all the CHs. As shown in Figure-2 a message with PDC and CH identifier is sent thru the backbone tree to the static sink, which is located at the corner of the network. Static sink maintains a prioritization table based on the CBPT algorithm.

Static sink is fed with information of optimum shortest distance from any cluster to all other clusters in the network. The shortest distance calculation is based upon Minimum spanning tree (MST) algorithm. The clusters are prioritized according to their PDC value. Cluster with highest PDC value is assigned to high priority followed by next highest PDC value. If more than one cluster possesses same PDC value then the cluster which is at minimum distance from the mobile sink's current will be selected as next destination to gather the data. Mobile sink receives the information about the path to which it traverses to gather the data from the static sink. Thus cluster that possesses data with high drastic changes the base station through mobile sink without latency.

CBPT Algorithm

Let PDC[i] be the value of PDC and PT[i] be the Priority table value for m clusters, Where i={1,2,….m}

For (i=0; i<m ;i++)

PT[i] = PDC[i]

For (i=0;i<m;i++)

For (j=i+1;j<m;j++)

If (PT[i]<PT[j]) then swap PT[i] and PT[j]

Route information obtained from MST routing table for each i^{th} cluster.

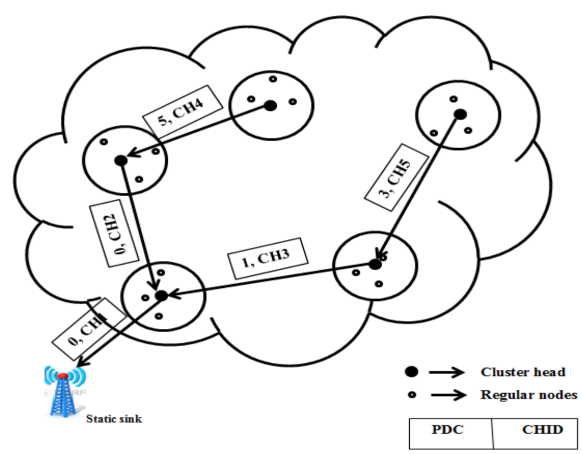

Figure-2. CH sending PDC message to static sink.

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Table-1. Routing data packet structure.

Priority	Route information
Starting	Intermediate nodes in the minimum distance from the starting point to cluster 4
CH 4	Intermediate nodes in the minimum distance from cluster 4 to cluster 5
CH 5	Intermediate nodes in the minimum distance from cluster 5 to cluster 3
CH 3	Intermediate nodes in the minimum distance from cluster 3 to cluster 1
CH 1	Intermediate nodes in the minimum distance from cluster 1 to cluster 2
CH 2	Intermediate nodes in the minimum distance from cluster 2 to finishing point

The Table-1 represents the simulated view of the routing information data packet structure for the example shown in Figure-2. The static sink frames the data packet by including all these routing information obtained from MST to the mobile sink.

The content aware nature of our algorithm makes the data collection highly reliable. The main purpose of this algorithm is to make, data with drastic change reachable to the base station without delay. The efficiency of this algorithm relies in the path traversed by the mobile sink.

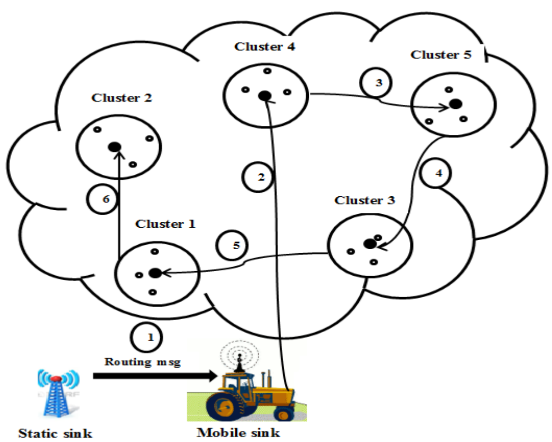

Figure-3. Traversing path of mobile sink.

7. SIMULATION AND RESULT ANALYSIS

In this section, we report our simulation results, each representing an averaged summary of more than 100 runs. More specifically, for each run of computation, the sensor field with or without an event region is generated using the procedure, the optimal threshold value is determined using the computation without event-driven data, and then revised centralized algorithm and CBPT algorithm is applied to obtain the values of the performance metrics. The averaged values of the performance metrics that are more than 100 runs are reported as our simulation findings. Simulation assumptions are mentioned in Table-2. The performance metrics include the message overhead and latency. In the setup the number of regular nodes considered are 100 and 4 cluster heads. Energy for regular nodes is 10.0 joules and energy for cluster head nodes is 100.0 joules. The transmitting and receiving power of regular nodes are 0.036 and 0.024 mwatts . The transmitting and receiving power of cluster head nodes are 0.3 and 0.6 mwatts respectively. Simulation time is set as 50 mseconds.

Table-2. Simulation assumptions.

Dimension	100m×100m
Deployment	Uniform
Sink location	Corner of the network
Location awareness	Provided

Figure-4 depicts that, for 100 nodes our revised algorithm shows a message overhead of only 138 bits whereas centralized algorithm shows message overhead of 176 bits. Our revised algorithm reduces message overhead upto 38 bits when compared to Centralized algorithm. Figure-5 depicts that, for simulation time of 150 ms our CBPT algorithm shows a delay of only 0.08ms whereas buffer based prioritization shows a delay of 1.5ms. Our CBPT algorithm reduces the delay of primary data upto 1.42 ms when compared to buffer based prioritization.

Figure-6 depicts that, for 250 nodes our CBPT algorithm delay of only 0.2ms whereas EDF prioritization shows a delay of 1.7ms. Our CBPT algorithm reduces the delay of primary data upto 1.5 ms when compared to EDF prioritization.

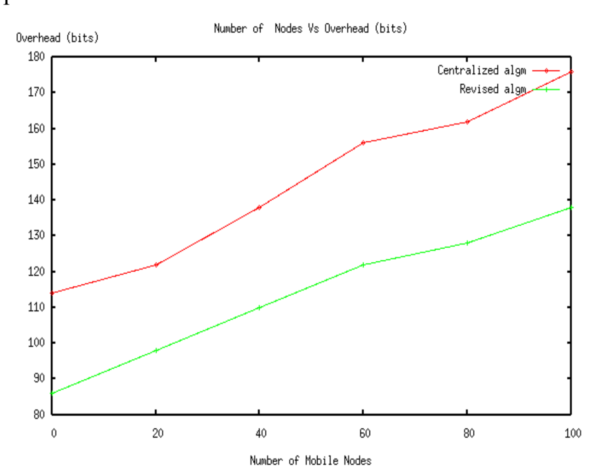

Figure-4. Message overhead decreased in revised centralized algorithm.

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

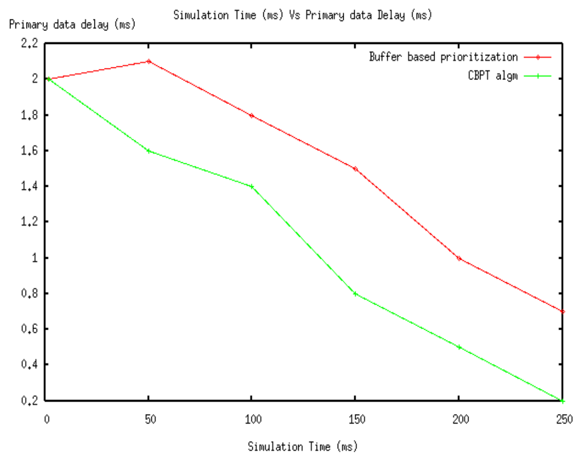

Figure-5. Delay comparision with CBPT and buffer based prioiritization algorithm.

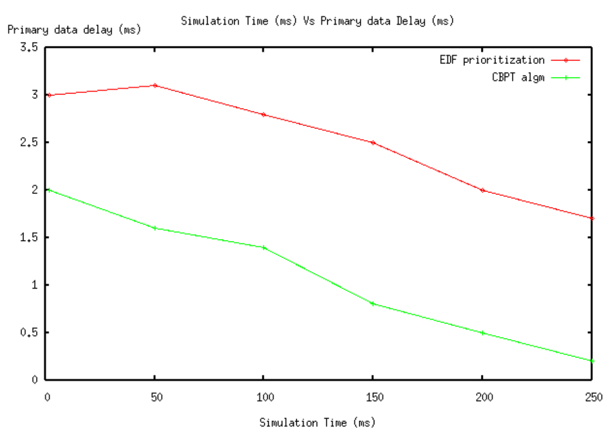

Figure-6. Delay comparision with CBPT and EDF algorithm.

8. CONCLUSIONS

Energy saving and delivering emergency data packets to the sink as soon as possible are the two important issues in disaster detection application. In this paper, our novel data-centric approach of collecting primary data without delay is a reliable method of data collection suitable for delay sensitive applications. Mobile sink in the network efficiently eliminates hotspot problem, which in turn enhances the network lifetime. Simulation results show that the CBPT algorithm guarantees the collection of primary data without delay. It proposes not only a method to detect important data based on the degree of change between periodically sensed data, but also provides different routing schemes according to the data importance. This approach with single mobile sink is applicable for small scale WSNs. Future work includes to extend the capability of our algorithm to support multiple mobile sink in a large scale WSNs.

REFERENCES

[1] Abdullah I. Alhasanat, Khaled D. Matrouk, Haitham A. Alasha'ary and Ziad A. Al-Qadi. 2014. “Connectivity-Based Data Gathering with Path-Constrained Mobile Sink in Wireless Sensor Networks”, Wireless Sensor Network, Vol. 6, pp. 118-128. Published Online, June.

[2] Ruitao Xie and Xiaohua Jia. 2014. “Transmission-Efficient clustering method for Wireless sensor networks using compressive sensing”,IEEE transactions on parallel and distributed systems, Vol. 25, No. 3, pp. 806-815, March.

[3] Can Tunca, Sinan Isik, M. Yunus Donmez, Cem Ersoy. 2014. “Distributed Mobile Sink Routing for WirelessSensor Networks: A Survey”, IEEE Communications Surveys & Tutorials, Vol. 16, No. 2, pp. 877-897.

[4] Arun A. Somasundara, Aditya Ramamoorthy, Mani B. Srivastava. 2012. “Mobile Element Scheduling for Efficient Data Collection in Wireless Sensor Networks with Dynamic Deadlines”Third International Conference on Computer and Communication Technology.

[5] Y. Yun and Y. Xia 2010. “Maximizing the lifetime of wireless sensor networks with mobile sink in delay-tolerant applications,” IEEE Trans. Mobile Comput., Vol. 9, No. 9, pp. 1308–1318, September.

[6] Hsu-Jung Liu1, Mei-Wen Huang, Wen Shyong Hsieh, Chenhuan Jack Jan and Ping Tung. 2009.” Priority-based Hybrid Protocol in Wireless Sensor Networks” 11th IEEE International Conference on High Performance Computing and Communications.

[7] Weili Wu, Xiuzhen Cheng, Min Ding, Kai Xing, Fang Liu and Ping Deng. 2007. “Localized Outlying and Boundary Data Detection in Sensor Networks”, IEEE transactions on knowledge and data engineering, Vol. 19, No. 8, August.

[8] Xu Ning and Christos G. Cassandras. 2006. “Dynamic Sleep Time Control in Event-Driven Wireless Sensor Networks”, Proceeding of the 45th Conference on Decision & Control, December.

[9] Byungrak Son, Yong-sork Her, and Jung-Gyu Kim. 2006.”A Design and Implementation of Forest-Fires Surveillance System based on Wireless Sensor Networks”, IJCSNS International Journal of Computer Science and Network Security, Vol.6 No.9B, September.

ARPN Journal of Engineering and Applied Sciences

[10] D. Chen, X. Cheng and M. Ding. 2004. "Localized Event Detection in Sensor Networks," manuscript.

[11] B. Krishnamachari and S. Iyengar. 2004. "Distributed Bayesian Algorithms for Fault-Tolerant Event Region Detection in Wireless Sensor Networks," IEEE Trans. Computers, Vol. 53, No. 3, pp. 241- 250, March.

[12] D. Li, K.D. Wong, Y.H. Hu, and A.M. Sayeed. 2002. "Detection, Classification, and Tracking of Targets," IEEE Signal Processing Magazine, Vol. 19, pp. 17-29, March.