ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 8(S9), 506–512, May 2015*
DOI: 10.17485/ijst/2015/v8iS9/69699

Compliant Data-centric Network Processing for Energy Economic Data Collection in Wireless Sensor Networks

E. Srie Vidhya Janani[1*] and P. Ganeshkumar[2]

[1]Department of Computer Science and Engineering, Anna University Regional Office Madurai, India; esrievidhya@gmail.com
[2]Department of Information Technology, K.L.N. College of Engineering, India; ganesh_me@yahoo.com

Abstract

In the recent years Wireless Sensor Networks (WSNs) have gained a significant attention from the research and industrial community. A real-time WSN is expected to perform exceptionally well with large-scale applications across the Internetwork. Most sensor nodes adopt un-rechargeable batteries with limited power supply. The practice of embedded and autonomous interrogation of data collected remotely across WSNs is one among the prime challenges faced in real-time environments. The process of data collection has a major impact in energy consumption across the sensor networks. The work proposes a Cluster-Based Priority Traverse (CBPT) technique where mobile sinks traverse in a dynamic path decided from the prioritized routing table. Also the work proposed emphasizes not only the collection of relevant data, but also a means of conditioning and interpreting these data. Sensed data are categorized into primary data that has a drastic variation from historical data and secondary data that resembles regular sensed data pattern. The WSN is organized as a clustered sensor network. The CBPT prioritizes the clusters in the network based on their Primary Data Count (PDC). The mobile sinks are ensured with high reliable traverse paths that include maximum high priority clusters to collect data in an energy efficient manner. A Minimum Spanning Tree is constructed for each cluster to all other clusters in the network. With repetitive simulations the categorization of data is achieved with high accuracy. Experimentations have been carried out to prove CBPT performs extremely well in terms of message overhead and delay.

Keywords: Cluster-based Priority Traverse (CBPT), Primary Data Count (PDC), Wireless Sensor Networks (WSNs)

1. Introduction

In the past decade, WSN has exposed an enterprising enhancement in disaster recovery applications. It plays an important role in forest fire detection[9]. There exist many researches to make fire detection more reliable. The environmental factors such as temperature and humidity sensed by sensors are not of high accuracy at times when there is even a small trace of smoke. In such scenarios, the mobile sink that traverses a static path around the network gathers data from the smoke originated region only when it gets into the transmission range of that particular node. Habitually treating all the sensed data with equal preferences may lead to delivery of event-driven data with higher latencies. The events of high interest are to be necessitates notifications without latencies. The sensory data is categorized as Primary data and Secondary data. The detection of primary data across WSNs is of much significance. The data instances that deviate from historical data patterns are termed to be primary data. On the other hand the sensory data with highest degree of spatial and temporal correlation is coined to be secondary data. The sensor nodes at closest proximities to the sink nodes are burdened with relaying a huge sum of data traffic from other nodes. This phenomenon is sometimes called the “crowded center effect” or the “Hot Spot problem”. The mobile sinks overcome the crowded center effect which is found to be impossible with static sinks. The mobile sinks migrate across a network as robots or vehicles on a prede-

Author for correspondence

termined path. Such scenarios emphasizes on network load distribution, continuous data collection continuously with the mobile sinks migrating intermittently at slower pace[5]. The proposed CBPT is a novel traversing technique allowing mobile sink nodes to follow the pre-determined data collection path. In turn it proves to be an efficient technique for the data collection across clustered sensor networks. The entire data collection relies on the prioritized table which well suits suitable delay sensitive real time applications.

In general, the network is divided into optimum number of clusters with optimum size[2]. The CPBT suggests the distributed implementation of centralized algorithm with slight revision, by rotating the Cluster Head (CH) role among the CH election candidates with high residual energy. Ultimately, a backbone tree is constructed with newly elected CHs. On the other hand, CHs with their energy drained out regain their energy by energy harvesting. Hence, the revised version of the algorithm reduces the message overhead of electing new CHs. Each CH monitors the sensory measurements across the network for data patterns that shows significant variance from the historical data patterns pre-recorded. This way the algorithm records primary data and secondary data across the clustered sensor network identifies those measurements that significantly deviate from the normal pattern of sensed data by computing the median of its member's sensed data, and then it finds the difference between individual sensed data and median.

The state of the art data collection approaches for large scale WSN environments is presented in Section-2. The need for clustering in WSNs is motivated in Section-3 and a briefing on the sensor network pattern is given. In the main body of the paper i.e. Section-4, the implementation methodologies for the proposed Revised Centralized Algorithm is discussed. The analysis of the primary and secondary data is discussed in Section-5. With Section 6, CBPT algorithm based on prioritized routing table is discussed. The implementation results are presented in Section-7 stating some general comparisons between the proposed and existing methodologies. Finally, the paper concludes with Section 8 outlining certain directions for future research.

2. Related Works

In the past, several research works have considered the opportunistic utilization of mobile sink. Many priority-based algorithms have considered buffer overflow[3] and time delay[1] as the primary metrics to prioritize the sensed data. Merely taking into account buffer size alone as a prioritizing metric is unaware of the significance of sensory data. The Early Deadline First (EDF) algorithm, involves time as a routing preference metric. The data with earliest deadline is termed with highest priorities without considering their spatial and temporal correlation. Hence energy efficient data communication becomes impossible with EDF. Such algorithms are found to be suitable for clock-driven applications than event-driven application. On the other hand, preference based data collection in a sensor network with mobile sink with static routing paths exposes higher latencies[4] and makes it unsuited for delay sensitive applications. A dynamic sleep control technique[8], for event-driven networks, does not provide any priority for traffic services[8]. The Priority based hybrid routing[6], proposes a new approach of geographic diffusion but limits with the utilization of only static sinks.

The event detection in terms of significant changes in sensory readings[10–12] is explored. The observations have been made with 0/1 decision logical computation. The related algorithms emphasize only on the most recently sensed data of individual sensors. The sensory measurements across the clusters are not exploited. The "change points" of the time series[10], are statistically computed. The event detector proposed[12] computes a running average and compares it with a pre-defined threshold, which can be adjusted by a false alarm rate.

3. Cluster Formation

Naturally, grouping sensor nodes into *clusters* has been widely adopted by the research community to satisfy the above scalability objective and generally achieve high energy efficiency and prolong network lifetime in large-scale WSN environments. The corresponding hierarchical routing and data gathering protocols imply cluster-based organization of the sensor nodes in order that data fusion and aggregation are possible, thus leading to significant energy savings. In the hierarchical network structure each cluster has a leader, which is also called the CH and usually performs the special tasks referred above (fusion and aggregation), and several common Sensor Nodes (SN) as members.

The cluster formation process eventually leads to a two-level hierarchy where the CH nodes form the higher level and the cluster-member nodes form the lower level. The sensor nodes periodically transmit their data to the corresponding CH nodes. The CH nodes aggregate the data (thus decreasing the total number of relayed packets) and transmit them to the Base Station (BS) either directly or through the intermediate communication with other CH nodes. However, because the CH nodes send all the time data to higher distances than the common (member) nodes, they naturally spend energy at higher rates. A common solution in order balance the energy consumption among all the network nodes is to periodically re-elect new CHs (thus rotating the CH role among all the nodes over time) in each cluster.

The BS is the data processing point for the data received from the sensor nodes, and where the data is accessed by the end user. It is generally considered fixed and at a far distance from the sensor nodes. The CH nodes actually act as gateways between the sensor nodes and the BS. The function of each CH, as already mentioned, is to perform common functions for all the nodes in the cluster, like aggregating the data before sending it to the BS. In some way, the CH is the sink for the cluster nodes, and the BS is the sink for the CHs. Moreover, this structure formed between the sensor nodes, the sink (CH), and the BS can be replicated as many times as it is needed, creating (if desired) multiple layers of the hierarchical WSN (multi-level cluster hierarchy).

In data gathering, clustering plays a significant role than routing tree structure. With clustering huge data traffic load becomes manageable together with optimal energy consumption. The network is organized with optimum number of optimum sized clusters[2]. The data flow of cluster selection algorithm is shown in Figure 1.

The nodes within *Hr* range are 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

The distributed implementation of centralized algorithm, states that, if a CH drains out of energy, new CH is selected by initiating the CH selection algorithm, which in turn increases the message complexity of the algorithm. The Revised algorithm claims that 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 with the inclusion of the new CH. Let EC = $\{C_1, C_2 \ldots C_i\}$ be the subset of election candidates, C_j be the next nearest node to the centroid of the cluster with high residual energy. In this case, C_j becomes the CH for the remaining rounds of transmission when CH_c runs out of energy.

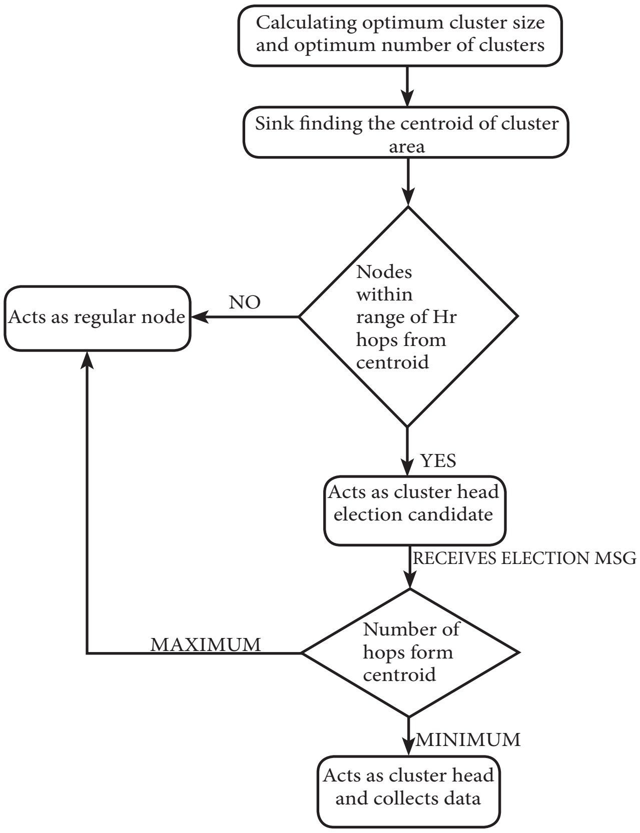

Figure 1. Data flow of CH selection algorithm.

4.1 Revised Algorithm

1 *If* $e(CH_c) = 0$
2 *And Let* $EC = e\,\{C1, C2,..Ci\}\,!= 0$
3 *Then* $C_j \in EC$, *next nearest node to centroid*
4 *With* $e(C_j) > \{e(E_C)\}$
5 $CH_c = EC_j$; *else Repeat 3 and 4 until* $j = i$
6 *End*

In connection to this algorithm j ranges from 1 to i. e (C_i) represents the energy level of the election candidate. C_{Hc} is the cluster head for current round of transmission.

5. Categorization of Data

The definition that distinguishes the outlier data and event-driven data is of prime importance. "An outlier or anomaly is an observation (or subset of observations) which appears to be inconsistent with the remaining data of that dataset"[7]. The outlier data collected in WSNs include noise and errors, actual events. The primary data detection provides an efficient way to search for sensory data that do not follow the historical data pattern of the sensor network.

The detected values consequently are treated as events indicating phenomenal changes in sensory data values that are of high 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. The sensor nodes are uniformly and independently deployed in a rectangular sensor field. The range of transmission of each sensor node is r. An assumption is made that 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 probability for occurrence of erroneous data within the network. The data with drastic deviation with historical data are clearly defined to be event-driven data.

Consider there are m clusters with n sensor nodes. Let *Ci* be the subset of clusters where i = {1, 2,…m} and *Sij* be the subset of sensor within a cluster i where j = {1,2,…n}. CH of the cluster *Ci* gathers the sensed measurements {*SM1*, *SM2,….SMn*} from its cluster members. CH arranges the measurements in an increasing order. Let *medi* be the median calculated from the measurements of cluster I, using *(1).*

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

The median proves to more accurate than mean because the drastic change a sensory data value experiences with its historical dataset cannot be clearly defined with mean. This is because the sample mean cannot represent the center of a sample, when some values of the sample are out of predefined range. However, median is an accurate estimator of the center of a sample.

Let *Dj* be difference between the *SMj* and *medi* as shown in *(2)*. It represents the difference between the sensed measurements of *jth* 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 is the standard deviation of the sample calculated with *(3)* and *(4)*.

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

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

Where *n* is the total of sensor nodes in the cluster. Standardize the data set to obtain *{X1,X2,….Xj,……Xn* with *(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 *ith* sensor is considered as primary data and the remaining data are secondary data. *TL* is threshold limit which is obtained from the test data compared with the learned predictive model for normal or abnormal classes. This predefined threshold is obtained from historical pattern.

6. CBPT Algorithm

The CBPT algorithm initiates with the Primary Data Count (PDC) from the CHs of the network. The CHs initialize PDC to be 0. It is 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 depicted 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. On the other hand, static sinks maintain a prioritization table based on the CBPT algorithm.

The static sink is fed with the information of optimum shortest distance from any cluster to all other clusters in the network. The shortest distance calculation is based on Minimum Spanning Tree (MST) algorithm. The clusters

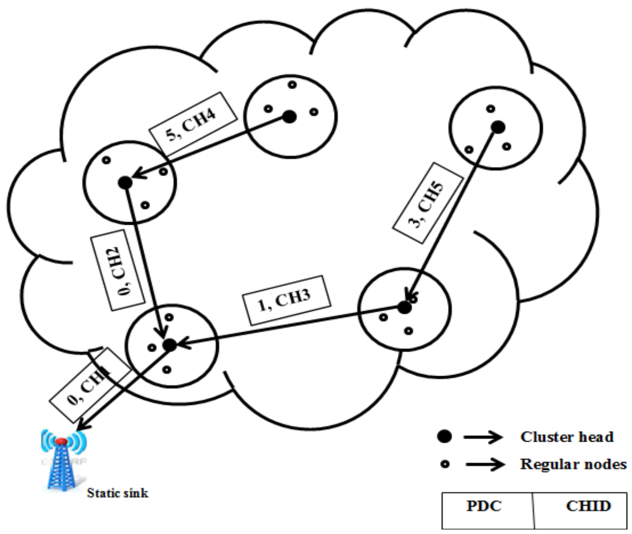

Figure 2. CH sending PDC message to static sink.

are prioritized according to their PDC value. The cluster with highest PDC value is selected as the next destination for the mobile sink to gather data. Likewise the clusters are organized across the path that will be ultimately traversed by the mobile sink for data collection. If more than one cluster possesses same PDC value, then the cluster which is at minimum distance from the mobile sink's current position will be selected as next destination to gather the data. Mobile sink receives the information about the path across which it traverses the sensor network to gather the data from the static sink. Thus, clusters that possess sensory data that drastically deviates the historical data communicates with the base station through mobile sink without latency.

6.1 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 ith cluster.

The prioritized routing structure for the illustration in Figure 2 is presented in Table 1. The static sink frames the data packet with the inclusion of 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

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

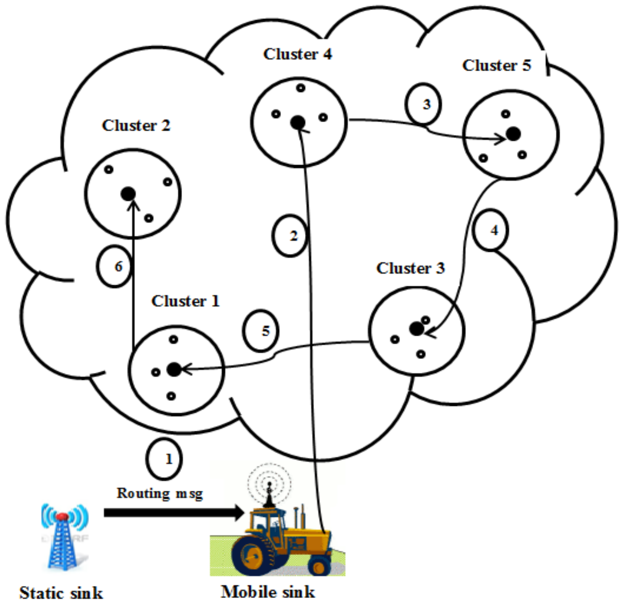

Figure 3. Traversing path of mobile sink.

to the base station without delay. The efficiency of this algorithm relies in the path traversed by the mobile sink.

7. Simulation and Performance 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 CBPT. The optimal threshold value is determined using the computation without event-driven data, and then revised centralized algorithm and CBPT algorithm is applied and their performances are proved with Quality of Service metrics. The performance measures are averaged with 100 runs of simulation. The experimental settings are mentioned in Table 2. The performance of the proposed work is measured in terms of message overhead and latency. The simulation is carried out with 100 regular sensor nodes and 4 CHs. The initial energy of regular sensor nodes is 10.0 joules. On the other hand, the initial

Table 2. Simulation assumptions

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

energy of CH nodes is 100.0 joules. The transmission and reception power of regular nodes are 0.036 and 0.024 mwatts whereas the transmission and reception power of CHs is 0.3 and 0.6 mwatts respectively. The simulation time is set as 50 mseconds.

With reference to Figure 4 the proposed revised algorithm allows 100 sensor nodes with message overhead of only 138 bits whereas centralized algorithm shows message overhead of 176 bits. Comparatively, it reduces the message overhead up to 38 bits.

Figure 5 for a simulation 150ms the proposed CBPT algorithm shows a delay of only 0.08ms whereas buffer

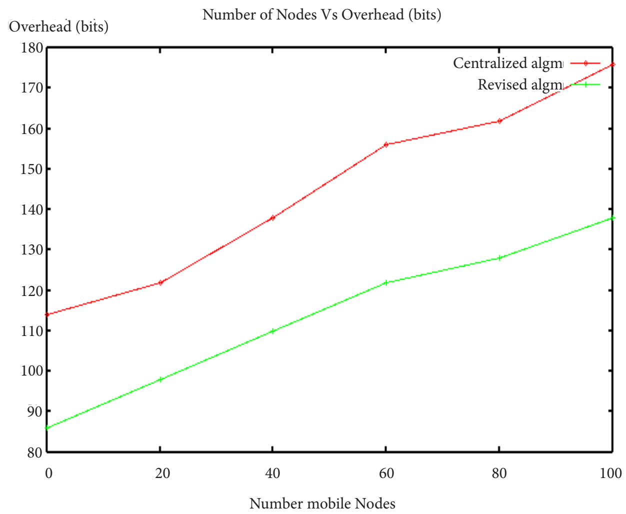

Figure 4. Message overhead decreased in revised centralized algorithm.

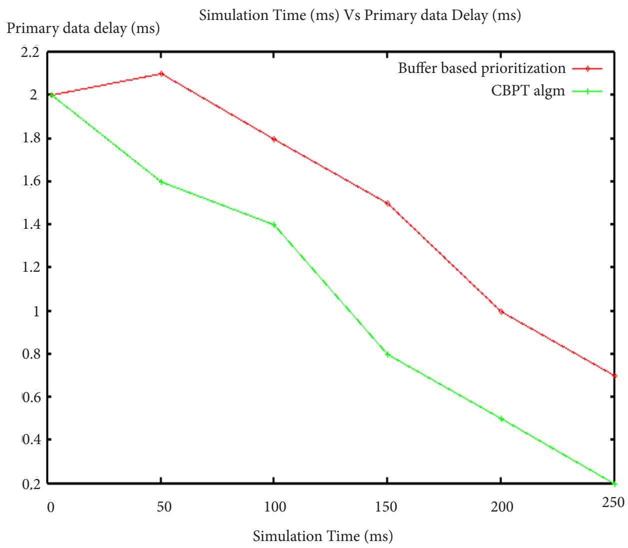

Figure 5. Delay comparision with CBPT and buffer based prioiritization algorithm.

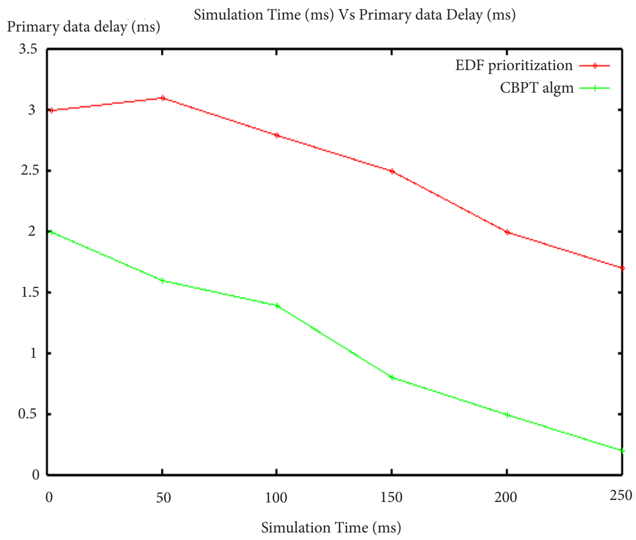

Figure 6. Delay comparision with CBPT and EDF algorithm.

based prioritization shows a delay of 1.5ms. Hence the CBPT algorithm reduces the delay in conveying the primary data by 1.42ms compared to the buffer based prioritization.

As illustrated in Figure 6 a total number of 250 nodes experience a delay of only 0.2ms with the CBPT algorithm whereas with the EDF prioritization a delay of 1.7ms is experienced. Hence the CBPT algorithm reduces the delay of primary data by 1.5ms compared to EDF prioritization.

The simulation results are presented. We compare our proposed algorithm CBPT with Buffer based prioritization and EDF prioritization algorithms. Also the Revised algorithm is compared with the Centralized algorithm. The performances are analyzed and the proposed work is proven to perform better in terms of message overhead and latency.

8. Conclusion

In time critical WSNs that support crucial applications such as disaster detection, the issues of prime importance are energy efficiency and emergency delivery of data packets to the sink with least latency. In this paper, our novel data-centric approach of collecting primary data without delay is proven to be one among the reliable methods of collecting data for delay sensitive real time applications. The Mobile sinks employed in the clustered sensor networks efficiently eliminates hotspot problem, which in turn enhances the network lifetime. The

CBPT algorithm guarantees the collection of primary data without any delay. It proposes not only a method to detect the significant data based on the degree of its actual deviation from periodically sensed data, but also provides prioritized routing schemes for routing the data across the sensor networks.

9. References

1. Alhasanat AI, Matrouk KD, Alasha'ary HA, Al-Qadi ZA. Connectivity-based data gathering with path-constrained mobile sink in wireless sensor networks. Wireless Sensor Network. 2014; 6(6):118–28.
2. Xie R, Jia X. Transmission-efficient clustering method for Wireless sensor networks using compressive sensing. IEEE transactions on parallel and distributed systems. 2014 March; 25(3):806–15.
3. Tunca C, Isik S, Donmez MY, Ersoy C. Distributed mobile sink routing for wireless sensor networks: a survey. IEEE Communications Surveys and Tutorials. 2014; 16(2):877–97.
4. Somasundara AA, Ramamoorthy A, Srivastava MB. Mobile element scheduling for efficient data collection in wireless sensor networks with dynamic deadlines. Third International Conference on Computer and Communication Technology; 2012.
5. Yun Y, Xia Y. Maximizing the lifetime of wireless sensor networks with mobile sink in delay-tolerant applications. IEEE Trans Mobile Compute. 2010 Sept. 9(9):1308–18.
6. Liu H-J, Huang M-W, Hsieh WS, Jan CJ, Tung P. Priority-based hybrid protocol in wireless sensor networks. 11th IEEE International Conference on High Performance Computing and Communications; 2009.
7. Wu W, Cheng X, Ding M, Xing K, Liu F, Deng P. Localized outlying and boundary data detection in sensor networks. IEEE transactions on knowledge and data engineering. 2007 Aug; 19(8):1145–57.
8. Ning X, Cassandras CG. Dynamic sleeptime control in event-driven wireless sensor networks. Proceeding of the 45th Conference on Decision and Control; 2006 Dec.
9. Son B, Her Y-S, Kim J-G. A design and implementation of forest-fires surveillance system based on wireless sensor networks. IJCSNS International Journal of Computer Science and Network Security. 2006 Sept; 6(9B):124–130.
10. Chen D, Cheng X, Ding M. Localized event detection in sensor networks [Manuscript]. 2004.
11. Krishnamachari B, Iyengar S. Distributed bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. IEEE Trans Computers. 2004 March; 53(3):241–50.
12. Li D, Wong KD, Hu YH, Sayeed AM. Detection, classification, and tracking of targets. IEEE Signal Processing Magazine. 2002 March; 19(2):17–29.