International Journal of Computer Sciences
Nov'2015 Issue
ISSN Number: 4142 3453

INTERACTIVE BIG DATA ANALYTICS FOR REAL TIME APPLICATIONS

Y. Vinaya Kumar[1], K. Rajasekhar Reddy[2]

[1]M. Tech (CSE)., Srinivasa Institute of Technology & Science, Kadapa, Andhra Pradesh

[2]Associate Professor, Head of the Department, CSE,Srinivasa Institute of Technology & Science,Kadapa, Andhra Pradesh.

ABSTRACT

Wireless Sensor Network (WSN) is an emerging technology. WSNs usually consist of a large number of small sensor nodes with limited onboard energy supply and deployed densely in a given area for information harvesting purposes. Since the sensor devices has limited memory and power capacity, the power consumption in WSN becomes as a major issue nowadays. So that, in the proposed framework, a scheme to reduce the power consumption in WSN is introduced. Proposed framework is clustering based. Clustering and Prediction techniques, which use temporal correlation among the sensor data, provide a chance for reducing the energy consumption of continuous sensor data collection. Thus it can achieve stability and prolongs network lifetime. An adaptive scheme is presented which is used to control prediction, analyze the performance trade off between reducing communication cost and prediction cost, and design algorithms to take the benefit of adaptive scheme to enable/disable prediction operations. Localized prediction scheme is performed which takes advantages over the previous dual-prediction scheme to minimize communication and computation cost thereby reducing the energy consumption. Sleep/awake scheduling can be applied. A practical algorithm designed for data aggregation will use faster and more efficient cluster-to-cluster propagation.

Keywords - *Wireless Sensor Network, Clustering, Adaptive Scheme, Prediction Operation, Sleep/Awake Scheduling*

1. INTRODUCTION

WSNs are composed of set of tiny sensor nodes, which can effectively monitor their surrounding environment. Due to the wide potential applications in battlefield surveillance, environmental monitoring, healthcare, weather forecasting, and disaster detection etc [1], WSNs have attracted quite attention from both academic and industrial fields in recent years. WSNs have a number of advantages over wired networks, such as simple, ease of deployment, extending transmission range, and self-organization. There are, however, a few inbuilt limitations to WSNs. These include small storage capacity, limited computation resources, low communication bandwidth, and limited device energy. In terms of energy, all nodes in a sensor network are battery-driven. Therefore reducing the energy consumption in sensor nodes and thereby increasing the network lifetime has become as major issue in WSNs.

In WSNs, hierarchical network structures have the advantage of providing scalable and energy efficient solutions. In this paper, different clustering algorithms are investigated and also compare these algorithms based on metrics such as cluster's load balancing, Cluster Head's (CH) selection, CH's role rotation, clusters overlapping, reliability, and location awareness.

In addition, energy-aware algorithms for reducing the energy consumption of sensors are discussed and designed. As such, one technique so-called prediction emerges to exploit the temporal correlation of sensor data. Cluster-based localized prediction technique is introduced which is highly energy efficient and simple over the previously studied dual-prediction technique due to the reduced length of routing path for transmitting data. An optimal sleep/wake scheduling algorithm is also used, which is helpful to attain a message capture probability threshold with minimum energy consumption.

2. BACKGROUND & RELATED WORK

Clustering techniques

The performance level of several clustering algorithms like LEACH, HCR, HEED, DECA, ASAP, PEGASIS etc., (which are designed with attention to energy-efficiency) is emphasized as follows.

- HEED clustering improves network lifetime over LEACH because LEACH randomly choose CHs, which may result in faster death of some nodes.
- CHs selected in HEED are well distributed in the network and communication cost is minimized as compared to other routing protocols.
- DECA has twice the energy efficiency than HEED in terms of CH residual energy.
- HCR, ASAP have considerable energy efficiency [1].

Prediction-based algorithms

Many approaches for energy-efficient monitoring have been explored. McConnell, Skillicorn [10] proposed that each sensor transmits to the base station (BS) the predicted target class rather than the entire raw data. Chu [3] proposed an approximate technique that uses prediction models to minimize communication from sensor nodes to the BS. Likewise, other data-driven processing to provide continuous data without continuous reporting was proposed. To do this, they developed a suppression strategy that adopts models for optimization of data collection. HEAR, HHEA energy aware routing algorithms are also analyzed.

Sleep / wake scheduling

Several researchers argue that TDMA protocols combined with sleep/wake scheduling are more suited to WSN applications (since TDMA protocols avoid energy waste due to contention). This requires precise synchronization between the sender and receiver, so that they can wake up at the same time to communicate. Most existing sleep/wake scheduling schemes assume that the underlying synchronization protocol can provide nearly perfect synchronization, and that clock disagreement is negligible. This, however, is untrue in practice. In the proposed framework, Sleep/wake scheduling of sensor nodes is based on the prediction operation.

3. PROPOSED FRAMEWORK

The proposed framework consists of following approaches to improve the energy efficiency: First design a Cluster and elect the CH in such a way to reduce the energy consumption, Second the adaptive scheme to enable or disable the prediction operation and sleep/ wake scheduling.

3.1 System Architecture and Design Issues

Since the performance of a energy aware protocol is closely related to the architectural model, this section strives to capture architectural issues and highlight their implications.

Network Dynamics: The proposed network architectures assume that sensor nodes are stationary. Dynamic events in the application require periodic reporting.

Node Deployment: This is application dependent. For our application the deterministic (square-grid) node deployment is chosen. In deterministic situations, the sensors are manually placed and data is routed through pre-determined paths. This will results in better average energy consumption and average coverage performance for the entire network area [12].

Energy Considerations: For deterministic node deployment (square-grid), the direct routing would perform well enough since all the nodes are placed in predefined location and are reachable to the CH. While scalability of the network is concerned, the multi-hop routing becomes unavoidable. Since the transmission power of a wireless radio is proportional to distance squared or even higher order in the existence of obstacles, multi-hop routing will consume less energy than direct communication.

Data Delivery Models: In the proposed model, event-driven data delivery model is used. The routing protocol is highly influenced by the data delivery model, mainly with regard to the minimization of energy consumption and route stability.

Node Capabilities: All sensor nodes are assumed to be homogenous. CHs are picked from the deployed sensors. It is assumed that the base station (BS) is more powerful than the sensor nodes in terms of energy, bandwidth and memory capacity.

Data Aggregation/Fusion: Sensor nodes consider the temporal correlation of data to send message to CH and CH performs data aggregation to send report to BS.

3.2 Clustering Approach

Clustering the sensor nodes as the basic of routing is an efficient mechanism for improving the energy efficiency. In clustering schemes, sensor nodes are dividing into a number of small clusters. Since deterministic node deployment is chosen (which is static) the cluster is formed based on the geographical area. It is essential to maintain the tradeoff between the number of nodes in each cluster and its communication cost, while forming the cluster. Each cluster has a coordinator/CH, and a number of cluster members. Cluster members can transmit data to their own CH directly, while CHs collect the data and send them to the sink node. Due to the following features clustering schemes are widely used in WSNs.

- Simple node coordination
- Redundant cluster nodes can be put into the sleep mode, since sensors within the sensing range of others have no need to be active all the time.
- They use multi-hop routing between CHs to avoid long-range transmissions.

Cluster Head Selection

The CH selection algorithm combines energy level of the nodes and average distance of the neighbors. Then on the basis of geographical area/ability to communicate with maximum number of nodes within a cluster, energy level of a node and least id, CHs are selected. Network deployment is considered as manual so the BS is well informed about the locations of the nodes.

Rotating the Role of Cluster Head

It is essential to rotate the role of CHs among nodes so as not to burden a few nodes with more duties than others and to easily power drained down. There are several possibilities for CH rotation. One way is to use a timer expiration to trigger the clustering algorithm. Another way is to use a dynamic parameter (e.g., remaining battery lifetime) for triggering the clustering algorithm at local regions. For example, a CH might trigger a new CH election process in its local region if its remaining battery lifetime goes below a prespecified threshold. It is obvious that more frequent CH rotation results in more clustering overhead and network interruption, while less frequent rotation may cause some nodes to die faster than others. The study of this trade-off is essential for achieving optimal network lifetime.

Cluster Formation Algorithm

Let us consider a WSN with 'n' stationary nodes deterministically deployed in a [X, Y] area.

// Topology setup phase

1) Specify Location; //Location of node in the WSN

2) Find Neighbors; //Check the neighbors of the

node and create a table

3) Find neighbor node distance; // Get the distance of

each neighbor and store it in the table.

// CH selection

4) For each node

If the node is alive // Energy> threshold nodes(k) having energy >=

average energy threshold.

If k >1 then

Case 1: //Deterministic deployment

Find out the nodes (n) with average

minimum distance from BS.

Case 2: //Random deployment

Find out the nodes(n) with maximum

number of neighbors.

If n > 1 // break the tie in CH election

Determine the node(x) with least id.

5) Assign node x as CH

6) Send CH announcement to all members, BS

// Rotate the role of CH

7) if timeout after m seconds

if energy level of CH < threshold then

Goto step 4;

3.3 Adaptive Scheme to Enable/Disable Prediction Operations

Prediction operation

In previous studies, the predictor training and prediction operations are carried out by the BS only, but not the sensor nodes, despite their increasing computing capacity. This solution while practical has many disadvantages, such as a high energy consumption incurred by transmitting the raw data to the BS, the need for wireless link bandwidth, and potential high latency. One solution which is provided in the proposed framework is clustering-based localized prediction [15]. It is highly energy efficient due to the reduced length of routing path for transmitting sensor data.

Adaptive Scheme

It is noted that unlike previous dual-prediction techniques, proposed prediction operation can be enabled/disabled. If the prediction operation is done at all times by the CH then it could increase the communication and computation overhead and thereby the network will consume more energy even though there is no any useful information. Therefore, it is designed to make the prediction by the CH only whenever it requires. i.e., the CH will make the prediction by using the temporal correlation among the information sent by the member nodes. As a result it says that if the correlation coefficient is too small, prediction will not be accurate. It is defined as follows.

If the local prediction is enabled then the sensor nodes follow the selective sending which is based on the ε-loss approximation: Given an error bound $\varepsilon > 0$, a sensor node sends its value xt to the CH if $|xt - xt'| > \varepsilon$, where xt' is a predicted representative data value to approximate the true data. That is, if a past value is close to the predicted value there is not much benefit by reporting it. Because the object may not move further or it will not be retained. If the value is much different from the predicted value, it will be send to CH by the sensed node. For that first a localized prediction model is developed.

3.4 Sleep / Awake Scheduling

A primary factor that prolongs the battery lifetime is allowing sensors to sleep when not active. This is due to the following three reasons.

- First, idle listening consumes significant energy that is comparable to transmission or reception. In contrast, the energy drainage during sleep time is about three orders of magnitude less than the reception energy consumption.
- Second, battery discharge is nonlinear, and some of the unusable charges can be restored in the battery after the sleeping period.
- Third, sensors are typically deployed redundantly, which implies that not all the nodes need to be awake simultaneously.

Therefore, if the application requires the sensors to continuously monitor the field for unexpected events, then a CH can determine which of its cluster members are redundant and advise them to turn themselves off. Thus, a CH maintains a minimal active set of nodes to cover the field in the cluster.

The variation with sleep/wake scheduling is based on the following observation. For some applications, the ε- approximation may not be strictly required. If the confidence level (of having data values within the ε- error bound) is very high, e.g., above a specified threshold, say αthreshold, the cluster members may never report data values to the CH. Therefore, there is no need for the cluster members to stay awake to obtain data values most of which will be discarded anyway. To allow sleep/wake scheduling for the cluster members, Lines (01)-(07) in Figure 1 is implemented, and by default, disable local prediction at cluster members. When a cluster member is awake, the CH checks if the member's data values are within the error bound with high probability. If yes, the CH will send a message to power off the member. The condition should be the confidence level αm is higher than the threshold αthreshold.

Periodic but infrequent collection of data from the cluster members is still necessary to perform accurate prediction. Let Δ be the time interval between two consecutive reporting by a member. We set the duration of a sleep period to mf*Δ, and when a cluster member wakes up, it will continuously perform data reading (and possibly reporting) for the next m*Δ time. Initially, mf is set to m. It can be increased if condition αm > αthreshold consistently holds, or decreased if the condition does not hold.

Algorithm at the cluster head
//sleep scheduling for members, Lines (01)-(07)
1. while member 'i' is awake
2. if timeout after m*Δ seconds
3. if condition αm > αthreshold holds
4. let member 'i' power off for mf * Δ seconds
5. while member 'i' is sleeping
6. if timeout after mf * Δ seconds
7. awake member 'i'

Fig.1 A variation with sleep/wake scheduling

4. PERFORMANCE EVALUATION

To evaluate the performance of our proposed framework, we conduct a series of experiments using NS2 simulator.

4.1 Simulation Environment

There are N nodes deterministically (square grid) deployed based on applications like battlefield or forest in WSN with certain range. All nodes will send their data with either direct or multi-hop transmission using the shortest path algorithm. The relevant simulation parameters are listed in Table I.

TABLE I

SIMULATION ENVIRONMENT

Parameter	**Value**
Network size	1250 X 900
Number of nodes	62
BS location	Inside
Data size(l)	2000 bits
Initial energy(E0)	2 J
d T	[250, 225]m
Δ	10 m

4.2 Cluster Model

The CH receives data selectively reported by all of its members, and performs local prediction on the sensor data. In this way, a CH can perceive an accurate view of all sensor data across the cluster, while communication cost is drastically reduced. The graph is plotted by comparing this proposed cluster-based, adaptive prediction and

International Journal of Computer Sciences
Nov'2015 Issue
ISSN Number: 4142 3453

sleep/awake scheduling scheme with non-cluster WSN architecture.

4.3 Average Energy Consumption

We compare the energy consumption with/without adaptive scheme to control local prediction. For that, only the energy consumption at all cluster members during the particular period is measured. On average, there are 4 clusters and 61 nodes. We output the sum of the energy consumption of all these 61 nodes.

4.4 Throughput

The performance of the algorithms is evaluated in terms of the total number of transmitted packets by all sensor nodes. For brevity, we only report the representative results. Proposed framework will yield the scalability because the distributed techniques perform data update and prediction locally.

4.5 Snapshot: Sample Scenario

Figure 2 shows the snapshot of our test result. In which at initial phase all 61 nodes are deterministically deployed using square grid node deployment. Then, at each cluster only sufficient number of nodes that are able to monitor the new object arrival from anywhere into the network area, are set to awaken state and rest of the nodes are set to sleep state. These awaken nodes are shown in purple color and sleeping nodes are shown in white color in the simulation. In addition, CH is always awakened. Hence only 5 nodes are awaked initially and rest of the 10 nodes are sleeping in each cluster.

It is designed to generate new object from any location within/outside of the area of interest and it can be randomly moved anywhere with random speed. When new object arrival/ new event occurrence is sensed by any one of the awakened node (which is shown in red color), immediately it will be reported to the CH. Then CH further reports it to BS and makes localized prediction. As a result, CH wakes up only sufficient number of nodes available in the predicted possible paths of an object movement and it is shown in purple colored nodes nearby red colored node. Rest of the nodes awakened already is sending to sleep mode and it is shown by changing the purple color into white. This is continuously progressed for the entire network as shown in Figure 3 and 4.

5. CONCLUSION

In WSNs, since the sensor nodes are energy constrained and have limited lifetime, energy consumption of sensor nodes becomes as a major issue. Though there are several work already done to reduce the energy consumption, still it does not attain the fruitful result and research is going on. Therefore, the goal of this proposed framework is to provide a better approach to reduce the energy consumption in WSNs and to prolong the network lifetime. It is achieved by two main approaches: 1) clustering-based: sensor nodes form clusters and elect the cluster heads in such a way to improve energy efficiency, and 2) prediction based: energy-aware prediction is used to find the subtle trade-off between communication and prediction cost. The detailed analysis and description of its two main components: adaptive scheme to enable/disable prediction operations and the sleep/ awake scheduling, is presented. Via performance evaluation, it is shown that it achieves energy efficiency even though the object arrived from any random location and moves randomly.

REFERENCES

[1] B.Baranidharan, B.Shanthi, AP/SASTRA University, "A Survey on Energy Efficient Protocols for Wireless Sensor Networks", IJCA, vol.11, Dec.2010.

[2] Bo-Chao Cheng, Hsi-Hsun Yeh, and Ping-Hai Hsu "Schedulability Analysis for Hard Network Lifetime Wireless Sensor Networks With High Energy First Clustering", IEEE Transactions on reliability.

[3] D. Chu, A. Deshpande, J.M. Hellerstein, and W. Hong, "Approximate Data Collection in WSNs Using Probabilistic Models," Proc.IEEE(ICDE), 2006.

[4] B. Gedik, L. Liu, and P.S. Yu, "ASAP: An Adaptive Sampling Approach to Data Collection in WSN," IEEE Trans. Parallel and Distributed Computing, vol. 18, no. 12, pp. 1766-1783, Dec. 2007.

[5] S. Goel and T. Imielinski, "Prediction-Based Monitoring in Sensor Networks: Taking Lessons from MPEG," ACM SIGCOMM Computer Comm. Rev., vol. 31, no. 5, pp. 82-98, 2001.

[6] Hongbo Jiang, Shudong Jin, Chonggang Wang, Senior Members, IEEE,"Prediction or Not? An Energy-Efficient Framework for Clustering-Based Data Collection in Wireless Sensor Networks", IEEE Transaction on Parallel and Distributed Systems, vol.22, No.6, June 2011.

[7] W.B. Heinzelman, A.P. Chandrakasan, and H. Bala krishnan, "An Application-Specific Protocol Architecture for WSN," IEEE Trans. Wireless Comm., vol. 1, no. 4, pp. 660- 670, Oct. 2002.

[8] Mahdi Lotfinezhad, Student Member, IEEE, Ben Liang, SeniorMember, IEEE, and Elvino S. Sousa, Senior Member, IEEE "Adaptive Cluster-Based Data Collection in Sensor Networks with Direct Sink Access ", IEEE Transactions.

[9] A.Meka and A.K. Singh, "Distributed Spatial Clustering in Sensor Networks," Proc. Int'l Conf. Extending Database Technology (EDBT), 2006.

[10] S.M. McConnell and D.B. Skillicorn, "A Distributed Approach for Prediction in Sensor Networks," Proc. SIAM Int'l Conf. Data Mining Workshop WSN, 2005.

[11] Sanjay Waware, Dr. Nisha Sarwade , Pallavi Gangurde "A Review of Power Efficient Hierarchical Routing Protocols in WSNs", IJERA,vol.2, Mar-Apr 2012.

[12] Wint Yi Poe, Jens B. Schmitt, "Node Deployment in Large Wireless Sensor Networks: Coverage, Energy Consumption,and Worst-Case Delay", AINTEC'09, November 18–20, 2009, Bangkok, Thailand.

[13] Y.Wu, S. Fahmy, and N. B. Shroff, ."Energy Efficient Sleep/Wake Scheduling for Time-Synchronized Multi-hop WSNs", Technical Report, 2006.

[14] Y.Wu, S.Fahmy, N.B.Shroff, ”Optimal sleep/wake scheduling for time-synchronized WSNs with qos guarantees”, in Proc. Of IEEE IWQoS, June 2006.

[15] Y. Xu and W.-C. Lee, “On Localized Prediction for Power Efficient Object Tracking in WSNs,”Proc. IEEE,(ICDCS), 2003.