DATA COLLECTION USING CSMA/CD AND EARLIEST DEADLINE FIRST (EDF) SCHEDULING IN WSN

Anand Deva Durai C  
College of Computer Science,  
King Khalid University, Abha,  
Kingdom of Saudi Arabia

Jeniffer J S C  
Dept of Computer Science and Engg,  
Aringar Anna College of Engg and Technology, Palani, India

Subash Chandra Bose  
College of Computer Science,  
King Khalid University, Abha,  
Kingdom of Saudi Arabia

Abstract- Wireless sensor network has an enormous scope of utilizations in detecting different sorts of parameters such as temperature, pressure, sound, pollution, etc. The sensed data in each sensor node are a valuable one. To reach the information to the base station for further processing, a lot of strategies are available. Each sensor senses the data in different sampling rate depending upon the sudden raise in the sensing parameters. The sensed data should reach the base station before the data becomes invalid due to the violation of the deadline.

In order to overcome the problem of deadline violation so that the sensed data becomes useless, this paper proposing a novel data collection algorithm based on the popular Earliest Deadline First (EDF) scheduling algorithm. This algorithm calculates the total utilization and start to collect the data by a mobile sink if the utilization bound is less than 100%. The various simulation parameters are taken into account to verify the performance of the proposed method and the result shows that it achieves high throughput, low delay, high Packet Delivery Ratio (PDR) and low energy consumption.

Keywords- Wireless Sensor Networks (WSN), data collection, mobile node, Earliest Deadline First (EDF) scheduling, deadline, Packet delivery ratio (PDR).

I. INTRODUCTION

Wireless sensor network (WSN) is sparsely deployed around the region to sense various details like enemy tracking in the army, weather forecasting, checking the quality of plants, monitoring patients’ health remotely when the doctor is not present, detecting and controlling city traffic congestion and can also be used to avoid road accidents and forest fires. The advantage of WSN makes it an important data collection agent. Each sensor node deployed in the sensing region works with limited battery operated power. The major concern next arising is making the data sensed by the sensors to reach at the destination point. Numerous data collection methods are available to efficiently gather the data from the sensor nodes and to deliver the collected data to the base station. Schemes through which data collection is done using the relay nodes are analyzed in [1], [2], [3], [4], [5]. The problem faced by the relay node based data collection method is that each sensor node will send their data by themselves or it will forward to the nearby nodes. Due to this more energy is depleted in these type of methods. To rectify the problem of relay node based data collection method, clustering based data collection methods were
proposed. The various approaches of cluster based data collection method are discussed in [6], [7], [8], [9], [10]. Fig.1a shows the clustering based data collection approach. Another approach for data collection is using mobile collectors to collect the data which are discussed in [11], [12], [13], [14]. In this approach a dedicated node called as mobile sink travels in a fixed or random path moves around the sensing region and collects the data from the sensors. Fig.1b shows the data collection approach using a mobile collector. The mobile collectors can work with single antennas as well as multiple antennas. The mobile sinks with multiple built-in antennas is using a data collection scheme called MIMO (Multiple Input Multiple Output). MIMO is achieving high efficiency in energy and time consumption. All the above mentioned data collection methods are focussing more on energy efficient and time efficient data collection, but not on delivering the gathered information to the base station within the stipulated time called deadline, which is more important in time critical realtime applications.

The major concern to be addressed here is to make the data to be delivered to the base station well before the deadline is getting violated, which will lead to disastrous results. The base station requires some data to be supplied to the destination urgently based upon the sudden changes occurring in the sampling rate. This has been addressed as a deadline violation problem in many researches in [24], [25], [26].

Deadline violation problem means when there is a sudden raise in the sampling rate of particular parameter gets increased, buffer in the corresponding sensors will be automatically get filled. The sensed data present in that buffer should reach the base station before it is getting destroyed or become useless. Many times it is very difficult to achieve because the base station or any mobile collecting agent shouldn’t be able to reach the data collecting sensor at the proper time.

The deadline violation problem leads to the data become useless. The research works addressing the deadline problem, have made use of a mobile sink to collect the data from all the sensors. This paper attempts to propose a scheme that eliminates the deadline violation problem by joining the the clustering schemes and the Earliest Deadline First (EDF) scheduling algorithm, which is more efficient in scheduling the realtime tasks without affecting the deadline. A mobile sink is a mobile node, which is assumed to have infinite buffer space and boundless energy, moves around the sensing region in a fixed path to collect the data sensors. The sensor nodes in the sensing region are grouped into number of square shaped clusters and the sensor node at the centre of every cluster and in one hop distance from other sensors in the cluster is selected as cluster head (CH).

The data sensed in every sensor is transmitted to the cluster head using Carrier Sense Multiple Access with Collision Detection (CSMA/CD)[28]. After collecting data from all the sensors, the cluster head will deliver the data to the base station without missing the deadline through the mobile sink which meets each cluster head in a predefined order fixed by the EDF algorithm.

The rest of the paper is organized as mentioned below. Section 2 discusses the related work. Section 3 presents the proposed framework to overcome the deadline violation problem. Section 4 provides the performance evaluation results. Finally section 5 concludes the paper.
II. RELATED WORK

The major issues to be taken care of in data collection in WSN are clustering, data collection at CH, sink mobility pattern and deadline preservation. Some of the existing works carried out related to these issues are discussed in this section.

Travelling Salesman Problem (TSP) is one of the classical and challenging optimization problems. It tries to identify a minimum time to cover all the nodes through the shortest path. The TSP has variety of applications in many fields and they are reviewed in [15].

TSP problem is NP-hard and is analyzed in [16]. Many heuristic based route construction algorithms have used the TSP algorithm in [17]. Various schemes used in TSP have been reviewed in [18] [19].

The Vehicle Routing Problem (VRP) can be stated as the method of constructing least cost route from one point to a set of other geographically distributed points [20]. Each node will have a service request based on their demand and each node will have a vehicle to route this request to the base station. Sometimes each node will have a constraint, based on which that node will visit in a particular time window and is explained in [21]. In the dynamic VRP explained in [22] the request to collect the data will come in dynamic manner depending upon the sampling rate of the sensor.

The deadline problem can be equated to the processor scheduling issue. Here the task will arrive in a periodic manner and it needs to be finished before the execution time gets over. The inter node travel time can be considered as the cost to travel between two nodes and service time of one node can be treated as the deadline for that particular node.

While serving one node there is a possibility that a request from another node may be generated simultaneously. This is given as the dynamic evolution of deadlines in realtime applications and is solved by Earliest Deadline First (EDF) scheme and is reviewed in [23].

The proposed work is employing a mobile sink which will travel and collect all the data in one round and deliver the collected data to the base station when it meets the base station. EDF scheduling algorithm is employed to create a schedule for mobile sink and the mobile sink meets and collects data from the CH one by one according to the schedule prescribed by EDF which preserves the deadline of every cluster head and deliver the collected data to the base station, which solves the deadline violation problem.

III. DEADLINE BASED DATA COLLECTION

This section describes the system model of the proposed Deadline Base Data Collection (DBaC) and formulates the Deadline violation problem.
A. System Model

The architecture of DBaC scheme is given in fig. 2. The sensing region is assumed to be in the shape of rectangle. The rectangular area of the sensing field is divided into grids of equal size. Each grid will contain a fixed number of sensors and also each grid is assumed as a cluster and a Cluster Head (CH) is elected in every cluster. The node which is located at the middle point of the grid is elected as CH. CSMA/CD scheduling scheme is employed by the CH to collect the data from the sensors within the cluster. Based on the deadline assigned to the sensor node in a cluster a priority level for each cluster is assigned and the mobile sink visits each cluster at a visiting point and collects the data from the CH. Popular processor scheduling algorithm Earliest Deadline First (EDF) scheduling is employed by the mobile sink to collect the data from the CHs. With EDF based on the priority, the mobile sink will calculate the total utilization using Earliest Deadline First (EDF) algorithm. If the total utilization is less than 100%, then collection of data according to the order of priority will not cause any missing of deadline. If there is a violation of deadline, choose the next set of priority to collect the data.

B. Problem Formulation

The proposed DBaC scheme solves the problem of invalid data data loss due to deadline violation and also the delay met by the packet to reach the base station. Let the WSN is considered as graph G(V,E) where the V is the set of sensor nodes and E is the set of links between the sensor nodes. An edge in the graph will exist if the sensor nodes are in the communication range. The end to end delay D faced by a data packet generated by a sensor node Vi to reach the Base Station is defined as $D(V_i, BS)$ in [28]. The end to end delay D is computed as in the equation (1) given below.

$$D(V_i, BS) = D_{int} + \sum dp \times k + T$$  \hspace{1cm} (1)

$D_{int}$ is the delay among the cluster members, $dp$ is the propagation delay, $k$ is the count of data packets generated and $T$ is the time consumed by mobile sink to reach the middle point of the grid.

The deadline $DL_i$ of the sensor node $Vi$ in a cluster is defined as the time duration into which the data sensed by the sensor node reaches the base station before it lost or becoming useless. Violation of deadline in reaching the sensed data to the base station is formulated as:

$$D(V_i, BS) < DL_i$$ \hspace{1cm} (2)

C. Data collection at cluster heads

a) Grid Formation

The sensing area is considered as rectangular region and is divided into number of grids of square in shape. Then deploy the total number of sensors equally in each square grid according to (3). No. of sensors $Cn$ in each grid based cluster $Ci$ is calculated as

$$Cn = \frac{N}{gn}$$ \hspace{1cm} (3)

where $N$ is the number of sensors to be deployed and $gn$ is the number of square shaped grids.

b) Dividing Rectangular Region Into Equal Square grids

Area of rectangle = $l \times b$ \hspace{1cm} (4)
Area of one square \( (A_{gi}) = \frac{(l*b)}{N} \) \hspace{1cm} (5)

Side of square \( (S_{gi}) = \sqrt{\frac{l \times b}{N}} \) \hspace{1cm} (6)

Number of squares along \( l = \frac{l}{s} \) \hspace{1cm} (7)

Number of squares along \( b = \frac{b}{s} \) \hspace{1cm} (8)

In order to divide the rectangular sensing region into equal number of square grids, measure the length \( l \) and breadth \( b \) of the rectangular region. Next fix the number of squares \( N \) that can be formed from the region. An area of one square is computed through (5).

Side of the square grid \( (S_{gi}) \) is the square root of the area of one square. Number of squares along the length and breadth can be computed using (7) and (8).

D. Cluster Head (CH) Election

The cluster head (CH) in each square shaped cluster is positioned at the mid point of the cluster to enable each sensor in the cluster to transmit the data packets in single hop. In order to determine the position of the cluster head, the centre point of a square shaped cluster region needs to be found out. The point at which the lines drawn from the centre of the sides of the square meets is considered as the centre of the squared cluster. 

\[(x,y) \text{ coordinates of square}\]

![Fig.3. Finding a center point of square using coordinates.](image)

Since each CH is situated in one hop distance, each sensor node will deplete almost similar energy. Only the sensors near the center point of the square grid can be elected as the cluster head.

Each sensor node will transfer the sensed data to the elected CH in the cluster for getting transferred to mobile sink. To avoid data loss occur at CH due to simultaneous reception of data from multiple sensor nodes Carrier Sense Multiple Access/Collision Detection (CSMA/CD)[28] the medium access and collision detection protocol is employed. The sending station checks whether any other station sending any data bits. If not, the sending station starts to transmit the data bits. Further, if any collision is not detected, the sending station will proceed with its transmission. Otherwise the sending sensor node will wait for a random time duration and terminate the transferring process if the random time expires.

CSMA/CD is used as a medium access control protocol in most of the LANs and early Ethernets. Any sending station will sense the physical medium for the presence of carrier voltage and detects whether any other station transmitting data. If so the sending station defers its transmission for a time quantum and tries again. This way of sensing the carrier avoids collisions between the transmission of stations and reduce the packet loss. The CSMA/CD is employed by deadline based data collection to avoid the collision occur due to simultaneous transmission of data from sensor nodes in a cluster to CH. The data collection scheme at CH using CSMA/CD is depicted in Algorithm 1.

The sensor node wanted to transmit the collected data to the base station checks whether the cluster head is busy in receiving data from any other sensor. The CSMA/CD procedure gets complete when the data are transmitted successfully to CH or getting failed if a collision is detected during simultaneous transmission of data.

**Algorithm 1**

If (Ready to transmit)

If (CH idle)

Transmit data to CH

Look for collision.

Else
Wait until CH is ready.
If (Collision)
End frame transmission.
Reset the retransmission counters.
Retransmit the data.

Algorithm 1: Data collection at CH using CSMA/CD

E. EDF scheduled mobile sink

EDF is a dynamic scheduling algorithm [17][18] used in various many real time applications to ensure that the task with nearest deadline is considered for execution. This method is adopted in Deadline Based Data Collection to create a schedule for the mobile sink to collect the data from all the CH in the predefined order. Whenever a sensor’s buffer gets filled, it will check whether its deadline closer than any other data collection process that is in execution. If so the data collection process with the closest deadline will be taken for execution next. Each arriving data collection process will have a particular deadline. So within the particular time the data need to reach the base station.

In order to ensure that all the deadlines are met till the completion of one round of mobile sink, it should calculate the total utilization factor U which is described in (9). EDF has the maximum utilization limit of 100%. The deadlines of all the sensors are assumed to be met when total utilization value is less than 100%. The utilization U is expressed in 3.7 as:

\[ U = \sum_{i=0}^{n} \frac{C_i}{T_i} \leq 1 \]  

(9)

where \( \{C_i\} \) are the worst case computation-time of n processes and the \( \{T_i\} \) are the respective deadlines of the data collection tasks.

Algorithm 2

Step 1: Calculate the absolute deadline (Time required to reach the data to BS).
Step 2: Generation of priority of clusters based on the deadline.

Calculate total utilization U using (9).

Step 3: If (U<100%)
Deadline of all tasks will meet, collect data from all grids based on priority order.

Else
Choose for another set of priority list that won’t cause missing a deadline by calculating the percentage of utilisation value

Step 4: Exit.

Algorithm 2: Scheduling the mobile sink using EDF

The detailed flow chart of the above mentioned scheme of data collection is given in the figure 4.

IV. PERFORMANCE EVALUATION

The simulation of the proposed EDF based data collection scheme DBaC is carried out and the results are compared with the popular existing data collection approach LEACH. Since the main focus of this paper is to meet the deadline of every sensor data and to reduce the loss of packets due to delayed submission to base station, the delay met in the packets to base station which is described in [1] is considered as the major parameter. However the results shows that it is achieved to a great extend. In order to compare the performance evaluation four parameters have been taken such as energy, delay, PDR and throughput. The results of comparison of the approach with LEACH protocol are shown in the figures 5 (a), (b), (c) and (d).

The proposed framework generates a cluster and the cluster head of this particular cluster is responsible for collecting and to transfer the data to the mobile sink. The mobile sink is travelling near to the cluster head to collect the data. While comparing to LEACH it’s an efficient energy saving method. In case of LEACH the cluster head will transmit the data to next cluster head, which is near to the base station. The cluster heads situating nearby base station loses their energy very soon. The graph in figure 5(a) shows that the energy
consumption with respect to the number of nodes. The energy consumption $E_{node}$ of a particular sensor can be calculated by $E_t + E_r + E_s + E_p$. The energy dissipation in a particular node can happen due to the following operation that it involves such as the energy spend in transmission ($E_t$), the energy spend in receiving ($E_r$), the energy spend in sensing ($E_s$) and the energy spend in sleep time ($E_p$). To compare the energy consumption $n=100$ nodes are considered. The proposed EDF based data collection method consumes lesser energy than LEACH. Delay means the time delay taken by the packet to reach the base station. If the data doesn’t reach the base station at a proper time it would result in useless data. If the important data are not reaching the base station, some negative consequences may arise. So reducing the delay is an important issue. LEACH depends on the neighbour cluster heads to forward data and that cluster head will also concentrate on sensing its own data and forwarding it to the base station. Obviously there are more chances for occurrence of the delay. EDF based data collection depends upon the priority the data to be collected by the mobile sink as soon as possible. Delay can be calculated by removing the receiving time component from sending time of a packet $i$. The total delay in the network is the sum of total delay and delay met by the $i$th packet. The total delay occurred at a LEACH protocol and EDF with $n=100$ nodes are given in figure 5(b). The comparison result shows that EDF has less delay comparing to LEACH protocol with respect to the increasing nodes in the network.
Fig. 5. Performance evaluation of EDF based data collection with LEACH. (a) Average energy consumption per node. (b) Data latency. (c) Packet Delivery Ratio rate. (d) Throughput rate.

Packet Delivery Ratio is the ratio between the number of packets generated at the source and the number of packets received at the base station. The packet delivery can be affected by many reasons in a network. The problem addressed in proposed framework is mainly concentrating on PDR. Deadline refers when the buffer of a particular sensor gets filled which means it will cause a buffer overflow. Due to buffer overflow problem there is more number of chances for loss of packets. In order to reduce the loss of packets due to buffer overflow, the proposed method uses a priority based data collection. This makes the sensors whose buffer getting filled soon will be collected immediately. Figure 5(c) shows the graphical representation of the PDR rate for both EDF and LEACH. It shows that EDF based data collection has a high packet delivery ratio comparing to LEACH.

The efficiency of a data collection method can be estimated by checking how many numbers of packets it can deliver to the base station within a particular amount of time. LEACH depends upon the cluster heads to relay the data they receive from other cluster heads and their own data to the base station. This scheme of data collection suffers from heavy network traffic. In case of EDF, a predefined time period is allotted to each cluster head to send its data and when a cluster head is transferring the data to the mobile sink the other cluster heads are excluded from sending their data. The throughput achieved by EDF based data collection is also high, comparing to LEACH as shown in Figure 5(d).

V. CONCLUSION

This paper, proposed an EDF based scheduling algorithm DBaC to collect the data from the sensors and to reach them to the base station before the deadline of the data getting violated. The mobile sink is used as a collecting agent which moves around the sensing region and collect the data from the cluster heads individually. The proposed scheme solves the problem of packet invalidation due to missing of deadline and delay occurring due to reach the packet to base station. The data collected by cluster heads are assigned a fixed priority based on the deadline of that data and based upon the priority the mobile sink will collect the data from the cluster heads. This novel scheme of data collection achieves maximum deadline saving, reduce the delay and packet loss due to data invalidation occurring in data collection in modern data collection algorithms to a greater extent.

REFERENCES


