An Efficient Data Redundancy Reduction Technique with Conjugative Sleep Scheduling for Sensed Data Aggregators in Sensor Networks

K. P. SAMPOORNAM
Electronics & Communication Engineering Department
K.S.R. College of Engineering, Tiruchengode, Tamilnadu
INDIA
sambu_swasti@yahoo.com

K. RAMESHWARAN
Electronics & Communication Engineering Department
Director, MAM group of Institutions, Tiruchirappalli, Tamilnadu
INDIA
krameshwaran@gmail.com

Abstract: - An Efficient Data Redundancy Reduction (EDRR) scheme extending the life time of sensor nodes on wireless sensor networks has been proposed in this paper. Normally, in Wireless Sensor Networks (WSN), more energy is needed for transmitting data. In our earlier work, conjugative sleep scheduling scheme with Connected Dominating Set (CDS) for parent selection has been proposed. Whenever a sensor node fails or a new node is added to the network, the conjugative scheduling scheme evaluates the route maintenance algorithm. Remaining power levels of nodes have been examined in order to increase the network lifetime. This proposed work integrates conjugative sleep scheduler scheme with data redundancy reduction scheme. This energy efficient redundancy reduction scheme basically utilizes Differential Pulse Code Modulation (DPCM) technique. Rake receiver model data aggregation is adopted for aggregating the data from various sensor nodes. Finally, the performance of EDRR scheme is compared with existing energy efficient protocols. Based on extensive numerical experiments, the proposed scheme is evaluated in terms of Throughput, Energy Drain Rate, Power Consumption and Network life time.

Key-words: - Sensor networks, Connected Dominating Set, Sleep Scheduling scheme, Data Aggregation, Redundancy Reduction, Power Consumption.

1 Introduction

In WSN applications, energy consumption has become a critical issue because of the impracticability of the replacement of battery. Sensor node spends its energy for sensing the environment, for processing the sensed information and for transmitting the processed data using transceiver. When the sensor node loses its energy, the network connectivity gets affected. Then, no packet can be transmitted to the base station.

Wireless Sensor Networks (WSNs) exhibits more traffic when all the sensor nodes are transmitting the data continuously. In addition, there exists a strong correlation between the data collected from the nodes deployed in the same target region. The nodes in a wireless sensor network forward the data towards a sink through other nodes. The restricted resources on the nodes need smallest amount of power to be depleted in this task. This demands a virtual backbone which can reduce the number of hops needed to reach the sink. To achieve this, the concept of Minimum Connected Dominating Set (MCDS) is used with the assumption that all nodes have identical transmission range.

A common scenario of sensor networks involves deployment of hundreds or thousands of low-cost, low power sensor nodes in the region from where information...
will be collected periodically. In sleep scheduling scheme, data distribution has been done by creating routing tree and data was collected by waking up the nodes in time and transmit data to user through the tree. The collected information can be further processed at the sink for end-user queries. In order to reduce the communication overhead and energy consumption of sensors while gathering, the received data can be combined to reduce message size [16]. To achieve this, a new data aggregation scheme was adopted in this paper. An important problem studied here is finding an energy efficient data aggregation scheme for gathering all data at the sink periodically in such a way as to increase the network life time. Since sensor nodes may generate significant redundant data, similar packets from multiple nodes can be aggregated. It will lead to the unnecessary power consumption. Data collected by many sensors in WSNs were typically based on common phenomena. Hence there is a high probability for data redundancy. Such redundancy needs to be exploited by a suitable redundancy removing technique which improves energy and bandwidth utilization.

In this paper we propose a new scheme called Efficient Data Redundancy Reduction (EDRR) scheme which integrates conjugate sleep scheduling scheme with data redundancy reduction scheme and hence reduces data communication overhead from the network. In addition to redundancy reduction, DPCM technique can also provide security to the network without using additional power resources, especially when we are dealing with high levels of redundancy.

2 Background Study

Newly, there is a huge increasing focus on smallest cost and low power consumption in wireless sensor networks. To achieve this, Power aware connected dominating set broadcasting approach presented in [19] is adopted in our scheme. In [1, 21], a distributed heuristic scheme was proposed for construction of energy-aware virtual backbone tree for WSN. A power aware MCDS construction was discussed in [3] which introduce the concept of threshold energy level for dominating nodes. An energy efficient MCDS construction presented in [20] considers the weight and the balance of energy. In paper [13] message overhead was minimized by modifying [20] with an efficient timer.

In sleep scheduling scheme, nodes sleep periodically; with a given sleep control policy [15], force a cost structure and look for an optimal policy in the middle of a class of policies. Thus, we think only on a single sensor node and focus on the tradeoffs between power consumption and packet delay [7]. We attempt to identify the manner in which the optimal sleep schedule varies with the length of the sleep period, the statistics of received packets and the charges assessed for packet delay and energy consumption [18].

In [8], the authors proposed a popular data aggregation paradigm for WSNs, called Directed Diffusion. It is a Data-Centric (DC) and application aware paradigm in the sense that all data generated by sensor nodes is named by attribute-value pairs. The main idea of the DC paradigm is to combine the data coming from different sources and reroute after eliminating redundancy. According to [4,17], network energy can be saved by minimizing the number of transmissions. Unlike traditional end-to-end routing, DC routing finds routes from multiple sources to a single destination that allows in-network consolidation of redundant data [7]. Aggregation scheme given in [10, 2], the path from each sensor node to the sink can be identified by a tree. The tree is enlarged by adding the shortest paths connecting the remaining nodes to the current tree. Another data aggregation approach called Tiny Aggregation (TAG) [11], in which simple queries are executed based on aggregation trees. This method utilizes other sensor’s state records to reconnect the lost nodes [5]. To remove the spatial correlations between sensed information, Redundancy Elimination for Accurate Data Aggregation in Wireless Sensor Networks (READA) implements grouping and compression mechanism [9].

Low Energy Adaptive Clustering Hierarchy (LEACH) [6] protocol is the first clustering protocol. It provides a conception of round. LEACH protocol runs with many rounds. Each round contains two states: cluster setup state and steady state. In cluster setup state, it forms cluster in self-adaptive mode; in steady state, it
transfers data. The time consumption of second state is usually longer than the time consumption of first state for saving the protocol payload. S. Lindsey et.al [10] proposed Power-Efficient Gathering in Sensor Information Systems (PEGASIS), a near optimal chain-based protocol which was an improvement over LEACH.

In [17], the authors proposed a localized, self organizing, robust and energy-efficient data aggregation tree approaches for sensor networks, which was called Localized Power-Efficient Data Aggregation Protocols (L-PEDAPs). They were based on topologies, such as LMST and RNG that approximated MST and efficiently computed using only position or distance information of one-hop neighbors.

### 3 EDRR Scheme with CDS Based Conjugative Sleep Scheduling

To extend the network lifetime, energy conservation is one of the most important constraint for WSNs because of the complexity in battery replacement. To accomplish this, short-hop communications are usually considered as a better choice when compared to the long-hop counterpart, mainly because of its low power consumption. However, due to lack of power management in sensors, power utilization mainly depends on the number of hops to be taken. More prominently, making fewer, longer hops permits nodes to expend longer in a low-power sleep mode, extending the overall network lifetime. Therefore, the number of nodes participated in routing to be significantly small. To navigate less number of hops in routing, we need to construct a wireless backbone to reduce the network size. In this work Connected Dominating Set acts as a backbone. Sleep Scheduling Scheme based on Connected Dominating Set is used in this proposed EDRR scheme to attain an excellent trade-off between power efficiency and Data accuracy.

#### 3.1 Conjugative Power Efficient Sleep Scheduling Scheme based on CDS

Conjugative sleep scheduling scheme [15] is employed to efficiently compute routing with localized energy levels in wireless sensor networks. It maintains better connectivity and satisfies a user defined coverage target conjugative to the sensor node and the sink. Sensor nodes use the residual energy levels and feedback from the sink for scheduling activity of neighbors. Sleep scheduler mode reduces randomness in scheduling and the sleeping nodes run in a synchronized schedule. It also maintains a backbone composed of active nodes which deliver data to sleeping nodes when they wake up and bound the communication delay between any two nodes in the order of one duty cycle.

The conjugate transpose matrix of an $\mathbf{a} \times \mathbf{b}$ matrix $\mathbf{X}$ with the complex entries is a $\mathbf{b} \times \mathbf{a}$ matrix $\mathbf{X}^T$. It is an iterative method and can be applied to methods that have enormous data to be handled manually. The resultant transpose matrix is applied with the complex conjugate for each entry. Here, we negate the imaginary parts without affecting the real parts. Defining the conjugate transpose as

$$\left( \mathbf{X}^T \right)_{mn} = \text{Negation}(X_{mn}) \quad \text{(1)}$$

In Equation (1) the subscripts denote the $(m,n)^{th}$ entry, for $1 \leq m \leq a$ and $1 \leq n \leq b$ and the negation denotes a scalar complex conjugate value.

Sensor node movement predicts future pickup points where the user expects a query result. In the case of pre-fetching, the sensor nodes send pre-fetched messages to future pickup points ahead of time. In sleep schedule mode, query dissemination is done for collector node which creates a routing tree, to alert the sleeping nodes. The data collection is carried out by waking up the nodes in time and send data to user through the tree. To minimize the energy consumption, a network is scheduled using Conjugate sleep scheduling algorithm which is described in the following section.

**Conjugative Power Efficient Sleep Scheduling Algorithm:**

- For each node $n$ in the tracking neighbors do
  - Compute the distance from the neighbors root node to node $n$
  - Compute $t_{\text{start}}$ and $t_{\text{end}}$
  - Compute the angle $\theta$ between the ‘root node ‘ and ‘node n’ and the instant velocity $v$
  - Compute DCmax.
  - Set the next cycle value as DCmax for node $n$. 

---

E-ISSN: 2224-2864 368 Issue 7, Volume 12, July 2013
• Set nodes duty cycle recovery number as Round (Tend – Tstart) / TC
• If Tstart = minimum_sleep_time then
  ➢ Set node n state as SLEEP.
  ➢ Reset node n waking up state as Tstart

3.1.1 Connected Dominating Set
A Connected Dominating Set (CDS) of a sensor network is a subset of nodes, known as relay nodes that create a connected network, such that any node in the unique network is either a member of the CDS or is within the transmission range of at least one node in the CDS. This article integrates power-aware connected dominating set based technique with data redundancy reduction for WSNs. This technique considers nodes that have a relatively greater energy level to construct CDS. For every round, a data gathering tree routed the data packet to the CDS Leader node, which is having more energy. This leader node will transmit the data to the sink. The non-CDS nodes are leaf nodes of the tree. Upstream node of an intermediary CDS node in the CDS data gathering tree is the neighboring CDS node. It is also relatively nearer to the CDS Leader. The resultant dominating set should include all intermediate nodes of any shortest path. In order to achieve this, the following algorithm is proposed.

Algorithm
• Create a sorted sequence of edges based on their transmission – reception energy loss.
• Randomly shuffle the sequence of sorted edges.
• Construct a spanning tree based on this shuffled sequence of edges.
• Remove the edges that are in the previously generated spanning tree(s) and obtain a reduced graph.
• Repeat the above procedure on the reduced graph and generate spanning trees until no more spanning trees can be generated.
• The internal node of a spanning trees form a Connected Dominating Set (CDS). Select the root node to be the node with the highest energy among the nodes in the CDS and form a rooted directed tree among the CDS nodes.

• The rooted directed tree along with the links from the leaf nodes to their nearest CDS node forms the directed graph tree.
• Repeat the data gathering process for every T rounds and then switch to the next edge disjoint spanning tree in the k-tree list.
• Compare this approach with just a MST- based DG tree – for every round. Choose the node with the highest energy as the root and form a rooted directed tree. The rooted directed tree along with the links from the leaf nodes to their nearest CDS node forms the MST- based DG tree.

In order to construct CDS, the following two rules are followed. Consider C and C’ are the connected and disconnected dominating set respectively. In C’, v and u are two disconnected vertices. In order to reduce the size of a Connected Dominating Set, Jie Wu et.al. proposed two rules [19]. For this, we first assign a distinct id, \( id(v) \), to each vertex \( v \) in C:
\[
N[v] = N(v) \cup \{v\}
\]
is a closed neighbor set of \( v \), as oppose to the open one \( N(v) \).

RULE 1: Consider two vertices \( v \) and \( u \) in C’. If \( N[v] \subset N[u] \) in C and \( id(v) < id(u) \), change the marker of \( v \) to F if node \( v \) is marked, i.e., C’ is changed to C’ - \( \{v\} \).

RULE2: Assume \( u \) and \( w \) are two marked neighbors of marked vertex \( v \) in C’. If \( N(v) \subseteq N(u) \cup N(w) \) in C and \( id(v) = \min\{id(v), id(u), id(w)\} \), then change the marker of \( v \) to F.

3.2 Power Efficient Data Redundancy Reduction
WSNs are characterized by high redundancy in the sensed data. Data fusion or Data aggregation is a solution to maintain robustness while decreasing redundancy in the data, but this mechanism also introduces latency and complications in QoS design. In WSNs security issue, data confidentiality, integrity and elimination of data redundancy are major requirements especially when the sensor network is deployed in a hostile environment.
Recently, many data aggregation protocols have been proposed with a focus on improving energy efficiency by eliminating the data redundancy in sensor data of the network. In this paper, to remove data redundancy, Differential pulse code modulation (DPCM) technique is adopted. Basic idea of DPCM transmitter and receiver is shown in Fig. 1(a) and 1(b). It is based on the fact that most source signals show significant correlation between successive samples. Since encoding uses redundancy in sample values, the bit rate is minimized. In this technique, signal received from the sensor node is sampled and then the difference between the actual sample value and its predicted value (predicted value is based on previous sample or samples) is quantized and then encoded forming a digital value.

![Fig. 1(a) Transmitter](image1)

![Fig. 1(b) Receiver](image2)

To combine the data from various sensor nodes Rake receiver model data aggregation scheme [14] is adopted as shown in Fig. 2. In order to separately detect M strongest multipath components, this receiver uses multiple correlators. Each correlator may be quantized of four bits. The outputs of each correlator are weighted to provide better estimate of the transmitted signal than is provided by a single component. According to the weighted outputs of the M correlators, demodulation and bit decisions are taken.

![Fig. 2 Rake Receiver model data aggregation scheme](image3)

This model detects M multipath components by using multiple correlators. The outputs are then weighted to provide linear combination of the correlator output. Each correlator detects a phase shifted version of the sensed data. The outputs of M correlators are denoted as $Z_1, Z_2, \ldots, Z_M$. They are weighted by $\alpha_1, \alpha_2, \ldots, \alpha_M$ respectively. Equation (2) describes the combined output $Z'$.

$$Z' = \sum_{m=1}^{M} \alpha_m Z_m$$  \hspace{1cm} (2)

Let $m_{x,y}$ be the total number of packets that node $x$ (a sensor) broadcasts to node $y$ (a sensor or base station) in a schedule $S$ with lifetime $L$ rounds. Since any valid schedule must respect the energy constraints at each sensor. For each sensor $x = 1, 2, \ldots, k$, the total energy consumption is given by the Equation (3).

$$\sum_{y=1}^{k} m_{x,y} A_{x,y} + \sum_{y=1}^{k} m_{x,y} TP_{x,y} + \sum_{y=1}^{k} m_{x,y} RP_{x,y}$$  \hspace{1cm} (3)

where TP is the transmitting power, $RP$ be the receiving power and $A$ is the power.
required for data aggregation. Each sensor, for all the $L$ rounds, makes one data packet that requests to be collected, probably aggregated and ultimately transmitted to the base station.

4 Performance Evaluation

The performance of EDRR is evaluated in which power consumption is considered to be a major constraint. To handle the topology alteration, a new CDS is to be constructed with low power consumption. The technique proposed here, will take care of node’s deletion from the network based on neighboring information.

In most of the real time applications, the data measured by the sensors are highly correlated either spatially or temporally. Data aggregation seems to be the commonly used communication technique which exploits data redundancy to dramatically decrease the amount of information to be transmitted thereby reducing communication traffic. It spends too much energy to transmit these redundant data. Since, data aggregation is introduced along with redundancy removing technique, it can reduce both network traffic and energy consumption. This reduced traffic will compensate the delay due to data aggregation. The parameters which are used in the simulation are Message Overhead, Energy Consumption and Network life time. In the primary part of the simulation, unit disk graphs corresponding to sensor networks were constructed.

Since the transmission power of a wireless radio is proportional to distance squared or even higher order in the presence of obstacles, multi-hop routing will consume less energy than direct communication. However, multi-hop routing introduces significant overhead for topology management and medium access control. Direct routing would perform well enough if all the nodes were very close to the sink. Sensor nodes are equipped with limited power source (<0.5 Ah 1.2V). But, node lifetime is strongly dependent on its battery lifetime.

5 Results and Discussion

For analysis, we considered a rectangular grid of points dispersed randomly over the grid. The size of the grid was chosen as per the number of nodes to make sure connectivity between the nodes. The nodes within one unit distance were chosen to be connected. In most of the cases, the node density was larger as per the wireless sensor network scenario. The EDRR and L-PEDAP protocols are run by using NS-2 simulator with 100-1000 nodes. Simulation parameters are given in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Size</td>
<td>1000 X 1000</td>
</tr>
<tr>
<td>MAC</td>
<td>IEEE 802.15.4</td>
</tr>
<tr>
<td>Simulation Time</td>
<td>900 s</td>
</tr>
<tr>
<td>Transmission Range</td>
<td>150m</td>
</tr>
<tr>
<td>Routing Protocol</td>
<td>CBQR (Cluster Based QoS Routing Protocol)</td>
</tr>
<tr>
<td>Traffic Source</td>
<td>FTP</td>
</tr>
<tr>
<td>Packet Size</td>
<td>150</td>
</tr>
<tr>
<td>Transmitted Power</td>
<td>0.660 w</td>
</tr>
<tr>
<td>Received Power</td>
<td>0.395 w</td>
</tr>
<tr>
<td>Idle Power</td>
<td>0.335 w</td>
</tr>
<tr>
<td>Initial Power</td>
<td>4.3 J</td>
</tr>
</tbody>
</table>

5.1 Performance Evaluation of Conjugative Sleep Scheduling Scheme

The experimental evaluation starts. First it is tried to choose the best parent selection strategy and then the experiments are continued with sleep scheduler strategy. The experiments with the defined topologies for this scenario with three parameters i.e., number of nodes $N$, maximum transmission radius $R$ and side-length of the square area $l$ is run. Another parameter defined in this topology is density $d$, which is defined as the average number of neighbors per node. Then a network with parameters $N = 80; R = 30m$ and $l = 120m$ is generated. The experiments are repeated for LMST, RNG and LMST with conjugate sleep schedule topologies. The life time obtained through these three methods is compared. This results show that (Fig.3) the LMST with conjugate sleep schedule strategy outperforms its alternatives in each case. The rest of the simulations evaluated the performance of the proposed routing scheme. Experiments were conducted with different values of $N, R$ and $l$. 
Similar to the first scenario, the degree plays an important role to determine the lifetime. So as in the first case, the maximum degree is increased slightly and the overall lifetime decreases. Fig. 4 shows that the performance of node life time with conjugative scheduler scheme is far better than other simulation topologies for increasing data nodes. As the area enlarges, connectivity decreases and distances get longer. It induces the sensed data time reduction.

To analyze the efficiency of the proposed technique, Throughput and Energy Drain Rate obtained through EDRR is compared with the existing L-PEDAP technique. By varying the number of nodes in a network from 100 to 700 throughput is measured. The results are tabulated in Table 2.

Table 2 No. of Nodes Vs Throughput

<table>
<thead>
<tr>
<th>No. of Nodes</th>
<th>Proposed EDRR</th>
<th>Existing L-PEDAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2.8001</td>
<td>2.2997</td>
</tr>
<tr>
<td>200</td>
<td>2.6593</td>
<td>2.0045</td>
</tr>
<tr>
<td>300</td>
<td>2.6071</td>
<td>1.9534</td>
</tr>
<tr>
<td>400</td>
<td>2.5592</td>
<td>1.8357</td>
</tr>
<tr>
<td>500</td>
<td>2.4591</td>
<td>1.6936</td>
</tr>
<tr>
<td>600</td>
<td>2.4912</td>
<td>1.5385</td>
</tr>
<tr>
<td>700</td>
<td>2.4892</td>
<td>1.3112</td>
</tr>
</tbody>
</table>

Fig. 5 shows that as the number of nodes increases, the throughput rate decreases in both the proposed EDRR model and the existing L-PEDAP model. But there is an appreciable improvement in the throughput obtained from the proposed EDRR model when compared to an existing L-PEDAP method.

Table 3 No. of Nodes Vs Energy Drain Rate

<table>
<thead>
<tr>
<th>No. of Nodes</th>
<th>Energy Drain Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed EDRR</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>100</td>
<td>1.284</td>
</tr>
<tr>
<td>200</td>
<td>1.4826</td>
</tr>
<tr>
<td>300</td>
<td>1.543</td>
</tr>
<tr>
<td>400</td>
<td>1.6937</td>
</tr>
<tr>
<td>500</td>
<td>2.6823</td>
</tr>
<tr>
<td>600</td>
<td>2.1948</td>
</tr>
<tr>
<td>700</td>
<td>2.4058</td>
</tr>
</tbody>
</table>
Fig. 6 is plotted to obtain the drain rate for the various number of nodes. From the graph, it is clear that as the number of nodes increases, the energy drain rate also gets increased in both EDRR and L-PEDAP method. The L-PEDAP method observes more energy rate for the processing of nodes in the network. When compared to an existing L-PEDAP method, the proposed EDRR method consumes less energy for the transmission of packet data from source to destination. If the energy consumption is less, the life span of the network will increase.

Fig. 6 No. of Nodes Vs Energy Drain Rate

To benchmark the proposed EDRR scheme, it is compared to some of the existing popular schemes such as L-PEDAP, PEGASIS and LEACH. Fig. 7 shows the total power consumption of the network with increasing numbers of source nodes. When the number of source nodes is greater, the aggregation tree is larger. It is shown that the EDRR scheme obtains the best solution quality when compared to existing three approaches.

Fig. 7 Performance Analysis in terms of Power Consumption

The lifetime of the network is nothing but the amount of time up to which the network can satisfy its coverage objective. From Fig. 8, it is clear that the EDRR scheme achieves the best network lifetime under all the settings. Although L-PEDAP scheme achieves almost the same network lifetime, the other two schemes have about 10 percent shorter lifetime than that of the EDRR algorithm.

Fig. 8 Performance Analysis in terms of Lifetime

Fig. 7 and Fig. 8 shows that the EDRR scheme gives better result than the L-PEDAP, PEGASIS and LEACH schemes, which means that the EDRR scheme is effective when a high percentage of nodes are sleeping.

6 Conclusion

The conjugative sleep scheduler scheme for data aggregation in wireless sensor network proposed an energy efficient routing approach. The properties of both minimum spanning tree and shortest path tree routing schemes are used. The most important contribution of this work is the availability of easily computable upper bound. The proposed mechanism uses the advantages of the powerful localized structures such as RNG and LMST and provides simple solutions to the known problems in route setup and maintenance because of its distributed nature and easily computable upper bound. Most of the applications designed for WSNs have high demands for efficient performance of the network. The integrity of data has tremendous effect on the performance for any such mission critical data gathering system.

In this paper, we have introduced an EDRR Scheme by applying the following
techniques: Connected Dominating Set broadcasting method, Conjugal Sleep Scheduling scheme, Rake receiver model data aggregation method and DPCM based redundancy reduction technique. Extending the lifetime of the network with continuous communication in the power constrained environment is one of the important issues directed in this work. The sleep scheduling technique is able to handle the small topology alters because of CDS node deactivation. This prevents the deep liberation of the battery and allocates it to refresh itself. This can expand the network lifetime. The accurate period of break and activity requires to be modeled to build an optimal local repair for EDRR heuristic to prolong a WSN’s life. The simulation results demonstrate that the EDRR efficiently extend network lifetime and balance node power consumption with a lesser message overhead when compared to the existing popular energy efficient techniques.

References


