Two pronged Strategy for Energy Optimization in WSNs by using In-network Compression and Synthesis of Multiple Queries at Base-Station

Vandana Jindal, A.K.Verma, Seema Bawa

Abstract— Wireless Sensor Networks (WSNs) find applications in environmental monitoring, healthcare monitoring, military surveillance, traffic monitoring etc. Immense data may be collected with the help of densely deployed sensors through the process in three steps - Data Acquisition, Data Processing and Data Communication. Desired information is extracted from this data by multiple queries. Sensor nodes of these networks are constrained in resources like energy and bandwidth. Due to high energy consumption in data communication major energy reduction works have focused on reducing this component only. In-network Data aggregation reduces quantity of data in communication for a simple query. Optimization of multiple queries reduces number of queries. In this paper query optimization at base station and in-network data compression techniques for achieving maximum benefit have been discussed. Both these, result into energy saving, thereby extending the life of node.

Keywords—Wireless Sensor Network, Multiple Query Optimization, MEMS, Epoch, SunSPOT, Solarium.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are made up of a large number of tiny sensor nodes spread randomly over a large geographical area. Miniaturization of sensor nodes with advances in Micro Electrical Mechanical Systems (MEMS) [11] technology has reduced the production costs of these nodes appreciably. Small size and affordable costs of nodes are the desirable features to enable their use in diverse applications like industry, science, transportation, civil infrastructure and security etc. To keep the size and cost at affordable levels these nodes are manufactured with limited resources like processing power, bandwidth and electrical energy. Power is consumed in every activity of these nodes i.e., sensing, computation and communication thus making it the fastest depleting resource of these nodes. As the sensors are deployed randomly in many inaccessible locations it is not possible to replenish the power. Once power of a node is depleted, it makes the node useless which affects the whole network badly. Therefore, primary concern is to use this resource judiciously. Cost in terms of energy consumption for sensing and computation is far less than energy consumed in communication. Therefore, besides optimizing sensing and computation activities, main thrust is on reduction in radio communication messages either through data compression in the network itself or through reduction in redundant queries at the base station or both.

In this paper, In-network compression and energy efficient multiple query optimization schemes are considered to reduce the quantity of data and the number of monitoring queries that are running in the sensor network. When a new query is submitted at the base station, the scheme checks whether any reduction is possible if the query is rewritten by synthesizing it with one of the already running queries through specifically designed algorithm. If so, the algorithm rewrites the newly injected query using the running queries. The rewritten query is then evaluated at the base station by making use of the results of currently running queries without being injected into the sensor network. As a result, the number of queries injected into the sensor network is reduced, resulting in lower energy consumption. Query rewriting is done at the base station therefore it does not affect the network, In-network aggregation/ compression is used by sensor nodes to reduce the quantum of data communicated to the base station for a particular query. Compression techniques, as the name suggests squeezes the data within the network and pass on just a small amount of the data to the sink.

The following section gives a brief summary of the works accomplished in related area. Sec III gives us a brief overview of energy as a constrained resource in WSN. Sec IV depicts the two pronged strategy for energy optimization used - In network compression and query optimization at the base station. Sec V describes the simulation setup for carrying out the proposed scheme and expresses the simulation results followed by the conclusion in Sec VI.

II. RELATED WORK

Cougar [2] and TinyDB [3] are the two most widely used database management systems to extract information from the database generated by sensor nodes. However, in these systems the focus has largely been on optimizing and executing a single long running query only. Unlike query processing in traditional databases, query processing in WSNs is different because of its own semantics, constraints and objectives etc. Same WSN is used for varying applications by multiple users therefore multiple queries pertaining to individual application are encountered in WSN. Handling of these multiple queries in an optimized fashion is area of research. Few studies and solutions for management of
multiple queries with a sensor proxy to control sampling rates have been proposed through Fjords architecture [4] proposed by Madden et al. and another proposal of SwissQM project [5] at ETHZurich. However, these proposals were successful in eliminating data redundancy to some extent and provided an approximate answer only (not the exact value). Algorithms too were proposed to optimize multiple region based queries. These were divided into partial aggregation sharing approach [6] and equivalence class approach [7, 8]. The approaches were not suitable for majority of the queries.

Energy efficient protocols have been devised. Low Energy Adaptive Clustering Hierarchy (LEACH) protocol proposed by Wendi B. Heinzelman et al. [9] is one of the widely used protocol. Many of its application specific variants [10-18] have been devised by the researchers. Work conducted on WSNs using data compression has also been the focus for quite some time now. Pradhan et al. [19] came up with the idea of distributed compression where source and channel coding were used. As a result the data transfer between the nodes due to compression was reduced. A distributed match source channel communication architecture and reconstruction method from noisy projections was proposed by Rabat et al. [20]. Similar gossip communication approach was proposed by Wagner et al. [21]. Although Wagner et al. [22, 23] came up with the architecture for distributed wavelet analysis eliminating the hypothesis about the grid regularity but failed to depict the method in selecting an optimal path for the compression and spatial correlation. Works on audio and video compression in WSNs have been found in [24, 25]. Efforts in improvising routing, aggregation, indexing and storage, energy balancing using compression could be seen in [21]. Sadilen and Martonosi came up with a new version of lossless LZW algorithm [26] which could compress 528 bytes data block. This S-LZW algorithm showed a significant amount of energy saving locally and globally, which employed buffering of data prior to its transmission. Comparisons of various compression codes updates for reconfiguring of nodes [27] was carried out by Tsiftes et al. presenting an algorithm with preprocessing and coding they experimentally showed 67% of energy saving using GZIP. Packet compression method (based on its frequency), was done by Ju and Cui [28] while they stated that packets with randomly changing fields were transferred uncompressed. Difference coding on Length variable coding were used for sensor readings. It was stated that 50% compression was permissible after removal of redundant data. The present work is an endeavor to obtain desired results in cost reduction through multi query optimization at base station and In-network compression. In this work the solution is sought by looking into base-station optimization in which data redundancy is removed followed by filtering of the result, ultimately moving towards decrease in the communication cost in terms of energy consumption by reducing the number of queries. Sensor nodes employ compression techniques thus a two pronged strategy to filter out redundant data has been devised. Taking into consideration the characteristics of WSNs and the constraints involved with respect to energy, our main objective is to minimize communication costs. Our work here presents the achievable energy savings when the sensor readings are compressed at the originating node.

III. MOTIVATION

Economic viability and long term reliability of WSN has always been the area of interest for research community. Every endeavor is to find ways and means which are frugal on energy consumption so as to increase the lifetime of the WSN. Energy is a major constraint in WSN. Therefore once energy is exhausted, node is dead which affects the network badly. The energy consumption involved for execution of a single instruction is 1nJ. However data transmission costs are manifold higher than data processing costs. 1Kb data transmission = processing three million instructions. Therefore to achieve maximum benefit in terms of energy consumption, it is essential that the number and content of transmissions is reduced without loss of other essentials such as data security, data fidelity and acceptable latency.

IV. PROPOSED TWO PRONGED STRATEGY FOR ENERGY OPTIMIZATION

Data is collected with the help of densely deployed sensors through steps of data acquisition, processing and communication. Initially when the Query set is empty, the query is injected into the WSN for data acquisition. The acquired data has to be communicated to the base station. The level of data compression achieved within the network, has a positive impact over the network’s energy consumption. Following depicts the two pronged strategy employed for optimizing the energy usage in a network.

A. In-network compression

There are many successive bytes within the input data stream, so we have selected compression algorithms which deliberately exploit these data structures. The other goal was the demand to achieve good compression gains while compiling to sizes within the resources available on current mote/ node platforms. Compression algorithms compared on Simulator were: Huffman and LZW. Besides the In-network compression results of which has been evaluated on simulator and is presented in the next section, base-station optimization as described below is also employed.

B. Base station Optimization

We are emphasizing on optimization of multiple queries at base station as this will help in reduction of energy consumption significantly. The queries are considered to be long running ones. In case of long running queries, mostly result dissemination messages are left in the network so a metric in terms of number of result dissemination messages in a unit time is considered to calculate or arrive at the cost involved in answering a query. Gain in cost reduction can be evaluated by finding the difference between the energy cost of newly injected query and synthesized query.

If we assume that each sensor node then has equal chance of being queried, selectivity of a predicate from the data is sel(p),
in case of multi hop query with sensor at depth d, cost of data acquisition query (Wq) shall be:
\[ W_q = \frac{\text{sel}(p).d}{s_i} \]
where \( s_i \) is sampling rate of the query.
Sel(p) depends on the distribution of attributes. Selectivity is computed over the whole data range. If a system has A attributes for querying and each attribute has a range which falls between \([\text{min}_i, \text{max}_i]\) (\( i = 1, 2, \ldots, A \)) the selection criteria of the predicate \( p \) having the range (upper value, lower value,) may be written as:
\[ \text{Sel}(p) = \sum_{i=1}^{A} \frac{\text{upper value}_i - \text{lower value}_i}{\text{max}_i - \text{min}_i} \]
If an existing query set \( Q_1 \) is synthesized into new query set \( Q_2 \) the cost difference is given as:
Cost difference = \( W_{Q_1} - W_{Q_2} \)
Where \( W_{Q_1} \) is cost of all the queries in \( Q_1 \) and \( W_{Q_2} \) is the cost of queries in new set \( Q_2 \). Value must be positive for synthesis to be beneficial. While synthesizing queries into new one it is to be ensured that new query is super set of queries being merged. Semantic correctness is also essential in case of data aggregation queries. Positive difference is denoted by a metric called Gain. Gain metric quantifies the saved cost in query rewriting. If we merge two queries \( q_1 \) and \( q_2 \) into one synthetic query \( q' \), it should be such that all the data requested by \( q_1 \) and \( q_2 \) must be requested by \( q' \).<br>
\[ \text{Gain}_{12} = \frac{\text{sel}(p_1)}{s_1} + \frac{\text{sel}(p_2)}{s_2} - \frac{\text{sel}(p_1 \cup p_2)}{\gcd(s_1, s_2)} \]
We shall write \( q_1 \) and \( q_2 \) into \( q' \) if and only if \( \text{Gain}_{12} > 0 \).
\( \text{Gain}_{12} > 0 \) only if
\[ \gcd(s_1, s_2) = s_1 \text{ or } \gcd(s_1, s_2) = s_2 \text{ Theorem[33]} \]
('s' refers to the sample period or the 'epoch')
In case Gain < 0, new queries are not integrated and go directly into the existing synthetic query list. If Gain > 0, queries are synthesized, new synthetic query is then checked for positive Gain with other queries of the set. The iterative algorithm is so designed that any achievable positive Gain is exploited fully. Due to generation and storage of data continuously by nodes in a sensor network a WSN may be considered similar to a distributed database [22, 29]. Assumption that the data is distributed in a database makes the usage of data more comfortable as modification of data becomes easy. Queries are injected to extract information from this database. Queries may be categorized as one shot queries and continuous queries. Queries reporting the current data only once are termed as one shot queries whereas continuous queries are those where the sensors produce and report the data periodically. Multiple queries being studied may be of any of these types.

C. Synthesis of Multiple Queries
The section here presents multi-query optimization algorithms. The base-station is the interface between the network and the user. User sends queries and obtains the result at the base station. Base-station is not resource constrained as the nodes. Therefore, base-station is used to filter out the redundant load of multiple queries into the network. Multi-query optimization algorithm rewrites a set of similar queries into a new set of queries before injecting them into the network, so that redundant data requests can be eliminated as much as possible. Correctness of semantics of queries is to be ensured while rewriting new query. All this is achieved with the help of different algorithm designed to obtain maximum Gain.

Treatment of a new query:
If a new query \( q_n \) arrives at a base station where results of a synthetic query set \( Q_s \) are already being calculated or obtained from the network, the algorithm will evaluate the benefit of rewriting the new query with the existing synthetic queries and find the most beneficial(in terms of cost) one. A new query is generated by merging the new query and existing query where the Gain is maximum. If there is no such query then new query is directly added to the set \( q_s \).
Iterative evaluation of cost reduction by integration of new synthetic query in the \( q_s \):
When a synthetic query is generated on arrival of new query as explained above the previously existing query set \( Q_s \) gets modified. It is evaluated through successive iterations through algorithm whether new synthetic query can be synthesized with any of the existing queries in \( Q_s \) to achieve further Gain. If so the pair is rewritten and the new one is again checked. This iterative process is continued till no further Gain is achievable.

When duration of a new injected query gets finished it is to be removed from the query set being evaluated. Query list is updated again to the previously optimized status prevailing before injection of the query.
In this way Multiple Query Optimization (MQO) design should be scalable-every new query is treated individually for positive Gain, energy efficient and adaptable-no redundant query should be there in the system.

Query Rewriting is employed for answering the newly injected queries by reusing the results of the already existing queries. This results into optimization of resource usage as duplicate data requests can be removed.
The notations used in the algorithm are:
\( Q = \{q_1, q_2, \ldots, q_n\} \) the set consisting of already running queries;
\( a_m = \) The attributes like temperature, pressure, etc;
\( q_{\text{new}} = \) a rewritten query of \( q_{\text{prev}} \) defined over \( q_1, q_2, \ldots, q_m \);
\( \text{EP} = \) Epoch is the time interval of taking the readings;
\( \text{sampling rate} \);
\( \text{SC}(q) = \) Selection Criteria of \( q \);
\( A(q) = \) Set of attributes listed;
\( S(q) = \) Set of attributes that are in selection criteria SC;
\( (s_c = p_1 \land p_2 \ldots \ldots) \) e.g.\( s_c = (\text{light} > 100) \land (\text{temp} \geq 30) \)
Decomposition of query \( q_{\text{new}} \) into \( d_{k1}, d_{k2}, \ldots, d_{kn} \).
\[ d_i = \prod_{\text{nodeid}} \sigma_{\text{sc}(\text{q}_{\text{new}}, a_i)(\text{sensors})}, \text{if } a_i \in A(\text{q}_{\text{new}}) \] 
\[ \prod_{\text{nodeid}} \sigma_{\text{sc}(\text{q}_{\text{new}}, a_i)(\text{sensors})}, \text{if } a_i \in S(\text{q}_{\text{new}}) - A(\text{q}_{\text{new}}) \]

For \( 1 \leq i \leq n \).

e.g. \( d_{\text{temp}} = \prod_{\text{nodeid}} \sigma_{\text{temp}(\text{q}_{\text{temp}} < 30)(\text{sensors})} \)
\( d_{\text{light}} = \prod_{\text{nodeid}} \sigma_{\text{light} < 100}(\text{sensors}) \)

**Algorithm for Rewriting the Query**

1. Let \( A(\text{q}_{\text{new}}) \cup S(\text{q}_{\text{new}}) = \{a_i\}; \) where \( i = 1, 2, 3, \ldots, n; \)

*/ composing the Prospect Query set*/

2. \( Q' = Q; \)

3. \( \text{for } j=1 \text{ to } m \)
   \( \text{for } q_j \in Q' \) do
   \( \text{if } (\text{EP}(\text{q}_{\text{new}}) \% \text{EP}(q_j) = 0) \land (\text{SC}(q_j) \land \text{SC}(\text{q}_{\text{new}})) = \text{false}) \)
   \( \text{remove } q \text{ from } Q'; \)
   end
   end

4. \( \text{for } i=0 \text{ to } n \) do
   \( Q' a_i = \{q | (q \in Q') \land (a_i \in P(q)); \}
   end

5. \( \text{for each } Q'a_i, (1 \leq i \leq n) \)
   \( \text{do compute test } C(\text{q}_{\text{new}}, a_i) \rightarrow V_{q \in Q'} a_i \text{ SC}(q) \)
   \( \text{if test == true} \)
   \( \text{hold;} \)
   \( \text{else} \)
   \( \text{return;} \)
   end

6. \( \text{for } i=1 \text{ to } n \) in \( A(\text{q}_{\text{new}}) \)
   \( d_{ai} = U_{q \in Q'} a_i \prod_{\text{nodeid}} \sigma_{\text{sc}(\text{q}_{\text{new}}, a_i)(q)}; \)
   end

7. \( \text{for } i=1 \text{ to } n \)
   \( S(\text{q}_{\text{new}}) - A(\text{q}_{\text{new}}) \)
   \( d_{ai} = U_{q \in Q'} a_i \prod_{\text{nodeid}} \sigma_{\text{sc}(\text{q}_{\text{new}}, a_i)(q)}; \)
   end

8. \( \text{Return } q'_{\text{new}} = d_{a_1} \otimes d_{a_2} \otimes \ldots \otimes d_{a_n}; \)

**V Simulation setup**

The goal of this simulation study is the evaluation of the achievable energy gains when data compression is applied prior to packet transmission. We examine a multi-hop scenario, where a node periodically delivers the data to the sink (base-station). The data collected through high density distributed WSNs are immense. In applications like temperature monitoring, the data collected at the free nodes needs to be transferred to the base-station periodically so that the data available is up-to-date. When readings are taken at regular time intervals, they are not expected to change significantly, but need to be transferred to the base-station at regular intervals.

SunSPOT mote used in the study is a WSN mote developed by Sun Microsystems. The device is built upon the IEEE 802.15.4 standard. Unlike other mote systems, The SPOT is built on Squawk VM (Virtual Machine) [30]. A SPOT is about the size of a 3×5 card with 32-bit ARM9 CPU, 1 MB RAM and 8 MB of Flash memory, a 2.4 GHz radio and a USB interface. The network platform of SPOT has built-in sensors along with the capability of interfacing with external devices. Two kinds of SPOTs i.e. free-range SPOT and base-station SPOT are present. The anatomy of free range SPOT has a battery processor board, a sensor board and a sunroof.

**VI TEST CASES, TEST RESULTS AND TEST ANALYSIS**

**A. Test Case I**

**Test Objectives**

The objective of the test case I is to verify that the functionality of one of the approaches i.e. in-network compression as proposed in the two pronged approach works in the direction of energy saving. The test is executed on the test bed “Solarium”. It has to be verified if the result obtained using the synthetic data fulfils our proposed approach. The compression algorithm used in Test Case-I is Huffman Compression.

**Test Results**

The final graph plotted between the number of queries executed against their respective sizes (in bits) i) without being subjected to compression and ii) with Huffman Compression is as under:
C. Test Case III

Test Objectives
The objective of the test case III is to verify that the functionality of one of the approaches i.e. base-station optimization proposed in the two pronged approach works in the direction of energy saving. The test is executed on the test bed “Solarium”. It has to be verified if the result obtained using the synthetic data fulfils our proposed approach.

Test Results
The graph in Figure 4 is obtained by employing three techniques i.e. i) firing independent queries, ii) query merging and iii) query rewriting along with merging of the already fired queries at the base-station. The values along the y-axis show the readings that were sensed and transmitted from the nodes to the base-station.

VII. CONCLUSION AND FUTURE SCOPE

In this paper we have proposed an energy efficient compression technique in-network to reduce traffic in WSN facilitating a better life time of the network. At the base-station query re-writing method is used for writing a new query if possible from the already injected queries. Both methods resulting into reduced energy consumption of the nodes. The work is implemented using the SPOT wireless platform. After the data has been sensed by the sensors of a free-range SPOT that is installed the data compression algorithm then comes into play. This compressed data in then transmitted to a base station SPOT which in turn sends them via USB to a PC. Even though various compression schemes are still under development, experimental results indicate that their compression rate and power reduction manners are quite
impressive. They are one possible mode to diminish resource constrain of wireless sensor nodes. On testing, the query optimization and processing algorithm described in this paper results in cutting the cost in terms of energy and bandwidth usage, both scarce resources in a WSN in sensing, computation and communication. Communication is the most energy intensive operation.

In the future as technology progresses the application areas of WSNs will become broader than the existing scenario. Their availability will become more to the masses than right now. There still exist many obstacles to be overcome for practical use of sensor networks

REFERENCES


Seema Bawa holds M.Tech (Computer Science) degree from IIT Kharagpur and Ph.D. from Thapar Institute of Engineering & Technology, Patiala. She is currently Professor in the department of Computer Science and Engineering at Thapar University, Patiala in Punjab (INDIA). Her areas of interest include Parallel and distributed computing, Grid computing, VLSI Testing and network management. Prof. Bawa is member of IEEE, ACM, Computer society of India and VLSI Society of India.