Modern Query Optimization Technique (Nature Inspired) for Improving Energy Efficient Data Gathering and Processing in Wireless Sensor Networks

Komalavalli Chakravarthi
Department of Information Technology
Jagan Institute of Management Studies
Rohini, New Delhi, India

Vijay Bhushan Aggarwal
Department of Information Technology
Jagan Institute of Management Studies
Rohini, New Delhi, India

ABSTRACT
One of the main obstacles in wireless sensor networks is limited energy and memory constraints. In wireless sensor network, communication among the nodes is the most energy consuming activity and needs to be minimized in terms of data size and number of communications. To overcome these obstacles the earlier research works introduced several techniques and methods for query optimization. Data aggregation, efficient routing, secured data transmission, secured communication, storage maintenance and various heuristics are some of the approaches introduced earlier. The main objective of this work is to accomplish better efficiency in terms of energy and data storage maintenance. To achieve this goal, a Modern Query Optimization [MQO] technique is proposed in this paper. MQO works at three levels namely (i) Monitoring the query and the path (ii) Tracking the data along with the node information and (iii) Choosing optimum nodes in a best path for data gathering based on query's lifetime. Artificial Immune System [6] algorithm along with the proposed three modifications is used to optimize the queries. The authors have simulated this modified algorithm in MatLab and the results show enhancement in efficiency.

Keywords
Wireless Sensor Network, Cluster Head, Energy, Artificial Immune System, Multiquery Optimization

1. INTRODUCTION
In wireless sensor networks many sensors are spatially distributed to monitor different phenomena based on the applications. These modern bi-directional networks enhance control of sensor networks. These are widely used in military surveillance systems, industrial applications and also in basic consumer applications. Wireless sensor network is a network which consists of large number of sensor nodes distributed in a geographical area for monitoring temperature, humidity, sound, pressure etc. Distributed control approach is mainly used in these types of networks to prevent failures and for gathering data in a better way. The data involved in the whole process is usually saved in the form of numerical data at cluster head and base station. These networks have their own design constraints. This includes processing and storage of data in each node as the storage capacity of sensor nodes is limited. Each node sends a huge amount of raw data to the central base station. Thus in order to improve the performance of the local sensor nodes their storage space is used as a distributed database where the queries are sent to retrieve the data. Many approaches have been suggested to manage the huge amount of sensed data. However they lack in completely enhancing the energy efficiency. Optimization of queries is the key to improving the efficiency.

Data management has to be highly efficient and distributed database approach paves a way for the best query optimization techniques. In this approach, the sensor nodes are treated as a data source, the sensed data being collected from different phenomena. The data is stored in the form of rows across a set of nodes [5]. Simultaneously the data collection should have the ability to minimize the energy consumption. Before transferring the data through distributed approach the data will be subjected to various data reduction techniques such as aggregation of data, merging of packets, compression and a process of fusion and approximation techniques. These data reduction can be performed at either at sensor node or cluster head. These techniques not only reduce the size of the data transferred, but also reduce the network load, communication tasks and the time for transmission. Thus storing and managing of the data in large amounts poses a very big challenge.

Research is being carried out around the world to improve the energy efficiency of data management by developing new data storage and querying methods. Distributed data approach has so far been viewed as the best method. Various query optimization techniques have been developed to further improve the energy efficiency.In the proposed technique, the data are not sent to the base station in a periodic manner. Rather they remain in the sensor nodes till queried. These queries are propagated in the network as per routing mechanism. On receiving the query, the sensor nodes send only the required data to the cluster head. This technique improves energy efficiency in data gathering in WSN, by optimizing the queries.

The rest of the paper is organized as follows. Section 2 includes Literature Survey. Section 3 describes monitoring the Queries and Section 4 describes the Query processing. Section 5 describes the proposed architecture and the algorithm for optimizing communication cost. Finally Section 6 gives Simulation settings, conclusions and future scope of this work. Section 7 gives references.

2. LITERATURE SURVEY
Several existing research works have discussed WSN and its applications, deployed for monitoring the environment and the natural elements. Sendra et al [10] discussed energy saving and optimization techniques by analyzing device hardware, transmission view, MAC and routing protocols points.

Vaidehi et al [7] proposed the algorithm "Distributed Nested Loop Join Processing (DNLJP)" here, since the limited memory and battery energy of sensors makes the network more challenging. This algorithm optimizes the query processed by the sensors in a distributed manner across all regions. This has an improved performance over a wide range of query.

For mining the data processed over the WSN a new behavioral pattern is introduced by Rashid et al [8] in their paper called "Share Frequent Sensor Patterns (SFSPs)". A growth mining tree method is suggested for all systems. To extract the useful information from the data collected by the sensors, the data management problem approach is used, which allows users to perform aggregation techniques on a sensor data such as MIN, COUNT, AVG. In-network aggregation systems are used to reduce communication cost due to the energy constraints.[15]

In centralized approach, data is stored in the base station and users are posting the query for retrieving the result. This was modified by making sensor nodes to act as local databases in a distributed manner. The data represents current state of an environment network. Real time database techniques use a distributed approach by considering the time constraints [12]. Data gathering techniques have been developed to save energy in wireless sensor network. The sensed data are collected using mobile collectors directly without any relay. Relay hop count provides better efficiency than the other gathering schemes [9].

The sensor network needs different approach for transient data in the continuous data stream. The data is summarized by the sliding window. It has fixed span width of elements of data. The hybrid histogram captures the changes in distributed data. This histogram supports the queries in the network [1]. A query engine and query operators are used to process the queries in WSNs. ADAGA (Adaptive Aggregation Algorithm for sensor networks) is proposed for query processing in the sensor nodes. This process dynamically adjusts the data collection and in sending the data to the cluster head and base station. It aggregates data and provides better efficiency [13] in the network. Most of the applications underlying wireless sensor networks are utilized to monitor the physical world by applying the query and analyzing the sensor data. In database storage data like temperature, sound, light etc. are stored. In multi query optimization, locations of the sensor node and type of sensor nodes are also stored for further processing [4] by recording the incoming query information.

From the above literature survey, it is clear that energy consumption during data transmission is a major problem in wireless sensor networks. Data gathering and aggregation is one of the main methods applied for reducing the energy consumption in WSN. But in case of high volume of data transmission during query processing in a distributed environment, it is necessary to consider various parameters such as energy, time, data transmission rate and cost. In this paper, MQO is proposed to provide a better solution for query optimization with focus on these parameters. This paper considers PDT – [Packet Driven Trajectories] that has certain distinct features of monitoring queries and can be exploited to achieve higher level of energy efficiency.

3. MONITORING THE QUERIES

In this paper, the authors propose a new query monitoring approach, which reduces the memory utilization and energy consumption, during inter node communications. MQO works by monitoring the queries, analyzing the queries and optimizing the queries. Initially monitoring the query is assigned as continuous job during collection of data. It always verifies the incoming requests for data collected in the network from various nodes and regions. This is in contrast to PDT algorithm, which monitors the query selectively by making the task selectivity aware [14]. Since WSN is spread over large area with huge number of nodes, it is essential to monitor the queries continuously, to revert back with the relevant answer to the appropriate request nodes. In order to save time and energy, the monitoring query process only monitors the necessary information about the requester, location of the request, direction and the path from source to sink, resulting in energy efficient process.

The earlier work had focused on the execution of a single long-running aggregation query. When new query enters it is not evaluated immediately. The query optimizer groups all the queries together and optimizes the queries group wise e.g. queries for same data. This approach uses a new model, which allows more number of users to post a sequence of queries. Moreover the proposed model completely eliminates the query preparation phase and passes the queries immediately for evaluation without saving in memory. As previously explained the monitoring queries phase stores the parameters like data, source node, location, destination node etc... for further process.

Then it is moved onto the processing section through the best node path. At processing scenario, the incoming queries are processed as soon as they reach the nodes. Thus this reduces the memory storage and time. The incoming queries are routed in such a way by checking the energy efficient nodal path and its reliability. By checking these parameters the overall performance will be increased.

4. QUERY PROCESSING

The algorithms presently executed in the nodes work in identical manner i.e. irrespective of energy level, query is processed in all nodes. Here the authors consider an algorithm which depends on the monitoring of the results contributed by each node. The first algorithm is referred as the Basic Algorithm which attempts in approximating the optimal cost and processing the query for the further results. This algorithm considers the node with set of queries. For a given rate, the incoming queries are monitored based on the conditioned parameters. For a large number of queries the most optimal reliable queries are chosen according to the algorithm. The node selects the approach that definitely yields least cost with greater efficiency. By processing these queries saves not only time and memory but also it prevents the nodes overlapped with accepted queries. When the efficiency seems to be higher at a node then the node will switch to its query process and sends up the result per updated query. Many research groups have focused at the concept of query processing in terms of energy consumption, frequency and time. The efficiency is taken as the important criteria in this query process.

Let a set of queries $\boldsymbol{Q} = \{\boldsymbol{q_1}, \boldsymbol{q_2}, \dots, \boldsymbol{q_n}\}$ come from L distinct users through base station. Let the sensor readings be represented as a set of vectors $\boldsymbol{v} = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\} \in \boldsymbol{\mathcal{R}^L}$. Every query $\boldsymbol{q_i}$ sends a request for a set of sensor data. This query information such as data requested, source node, location, destination node are stored in table $\boldsymbol{T}$ for further comparison. Then $\boldsymbol{q_i}$ is evaluated by comparing the vector data $\boldsymbol{v_i}$.The matching query $\boldsymbol{q_i}$ is then processed, while the rest are eliminated. During this query processing time, the data is recorded in T in order to check the path, length of the path, sending and receiving node availability and trustiness of the nodes. From this, according to the shortest path, node availability and trustable-node the $\boldsymbol{q_i}$ is processed. Node trustiness can be calculated by computing the nodes behavior and verifying the node information i.e., by evaluating parameters like whether the node belongs to the same network, in time-response, in-time acknowledging to other nodes.In case of multi query optimization it is aimed to reduce the cost of query evaluation alone while ignoring the computation issues during communication [11]. In the sensor networks, the fully distributed algorithm considers the memory constraints and computation. Let us consider queries as $Q = \{q_1, q_2, \ldots q_m\}$ over the κ distinct sensor data sources. The sensor readings set is a vector denoted as $x = \{x_1, x_2, \ldots x_k\}$ which belongs to $\Re^k$.

Here, each query q_i requests a combined value from the subset with desired frequency where each query is expressed as a κ-bit vector. When x_j contributes to the value q_i, then the vector element j is 1 or else it is 0. The dot product (.) of query and sensor reading estimates the value. It is represented as $q_i.x$. The combinations are executed in a multi- query optimization when there is a provision of dissemination tree which connects κ sensor nodes and the gateway. The aim of the approach is to minimize the communication cost by devising a plan to evaluate queries.

5. OPTIMIZING THE QUERIES

When a query enters into MQO, the monitored data is investigated and optimized using Artificial Immune System. AIS is one of the best optimization algorithms used in various kinds of applications. Artificial Immune System mimics the human Immune System in solving problems by using distributed adaptive systems.

The Clonal Selection model of AIS consists of affinity computation with mutation based solution finding, which helps to solve any optimization issues. This cloning method is the fundamental principle of the clonal selection methodology, where newly cloned cells will be mutated to maximize the affinity value of the antibodies. This process provides progressively better response to neutralize and eliminate the antigen. Applying this methodology, the number of solutions after mutation is proportional to the affinity value.

5.1 AIS algorithm with MQO[3]

1. Initialize a random population P1
2. Compute Optimum Final Value[energy] for each generated population P1
3. The affinity value is calculated by 1/OFV for P1 where OFV is Optimum Final Value
4. Compute the rate of cloning $(ROC = (\frac{1}{OFV} * P1)/\, Total\ affinity\ value)$
5. Generate the clones of the problem, according to the ROC.
6. According to the Cloning the size, the number of population is increasing into S1
7. Apply both inverse and pairwise mutation on S1.
8. Arrange all the population in ascending order and maintain the original population size [P1] by eliminating the highest OFV based clones.
9. Replace R% of the highest OFV based solutions by new random population generation.
10. Repeat (2) until obtaining a best OFV.

Entire functionality of the AIS algorithm with MQO has been described in terms of algorithms, Pseudo code for optimizing the query processing in WSN. To obtain the OFV, the size of population (P1), number of iteration K1 and replacing factor (R1) are, changed from a minimum value into a maximum value. In this experiment,

P1 = 50, 100

K1 = 100, 500, 1000

R1 = 10%, 20%

The above values are assigned and used. Even though, the number of population and iterations are changed between the minimum values into the maximum value, the objective value is obtained at different timings for different number of queries. Especially, the objective value is obtained when P1=100, K1=1000 and R1=20%.

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For three authors, you may have to improvise.

5.2 AIS - Pseudo Code

1. Let us assume P1 is the randomly generated population
2. Assume E_n is the OFV for each population P1
3. Let K1 be the number of iterations
4. For i=1 to K1
5. For j=1 to P1
6. Affinity-Value(j) = 1/OFV
7. End
8. For j=1 to P1
9. ROC (j) = (1/OFV * P1) / $Total\ affinity\ value$
10. S1 = Generate clones for each ROC(j)
11. End j
12. Mutate inversely S1
13. Compute OFV for S1
14. If (OFV(S1) < E_n) then E_n = OFV(S1)
15. Mutate pairwise S1
16. Compute OFV for S1
17. If (OFV(S1) < E_n) then E_n = OFV(S1)
18. Arrange S1 in ascending order and take only the top P size clones from S1.
19. Then, to improve the original P1, replace R% of highest OFV based clones from S1 by R% of newly generated population to maintain the original size of P1
20. End I
21. Finally $\mathbf{E_n}$ is taken as the objective value for the entire problem

Let Qn denotes the Query number, Sn denotes the source node, Dn denotes the destination node, pd denotes the path distance and ntv denotes the node trust value.

P1 = 10

K1 = 50

R1 = 10%

Number of queries = 10

String S1 ={ Qn, Sn,, Dn,, $\boldsymbol{pd}$, $\boldsymbol{ntv}$,En } and S(i)={Qn (i), Sn(i), Dn(i), $\boldsymbol{pd}(\mathbf{i})$, $\boldsymbol{ntv}(\mathbf{i})$,En(i) } best string $\boldsymbol{S}$ can be chosen by applying the below equation(String represents the set of all parameters to be evaluated or used for optimization process)

$$\boldsymbol{Qd} = \begin{Bmatrix} \mathbf{min}(\boldsymbol{pd}(\boldsymbol{i}) \\ \mathbf{max}(\boldsymbol{ntv}(\boldsymbol{i})) \\ \mathbf{min}(\boldsymbol{En}(\boldsymbol{i})) \end{Bmatrix}$$

Qd is the optimized query having minimum path distance, maximum node trust value and minimum energy.

6. SIMULATION SETTINGS

To experimentally verify the proposed AIS with MQO algorithm, MATLAB software is used and the time, cost and energy are calculated for processing each query qi. The q_i is tested using Artificial Immune System algorithm and optimize the q_i by time, nodes, shortest path. Time and the distance ultimately help to reduce the cost of the data transmission. We also show the benefits of the MQO algorithm by drawing real time data taken from a real sensor network infrastructure deployed in the Intel Berkeley Research Lab. MQO aims to reduce the energy consumption by optimizing the query process and it is taken as the primary objective in communication protocol designed for WSN. Several MAC and routing protocols have been proposed for energy consumption. But this study focuses on query optimization based energy reduction in WSN. The simulation settings used for experimenting MQO is given in below table-1.

Table 1: Simulation Settings

Parameters	**Values**
Network Size	1500 x 1500
Number of Nodes	200, 500, 1000 and 2000
Number of Queries	100, 200, 500 and 1000
Data size	10 bytes, 15 bytes and 20 bytes
DB	Tiny DB
Simulation Time	50 Sec
Number of Rounds	10
Number of Optimized Queries selected in each round	10 to 50
MAC	802.11 and above
Network	All kind of WSNs

It is essential to assign certain values for simulation parameters to obtain best results. The simulation parameter and the relevant values are given in Table-1. The simulation is carried out 10 times [considered as rounds] with different values assigned for the parameters. Sometimes the nodes are considered as moving nodes. The results obtained in each round are given below with detailed description. From the results shown in Figure-2 to Figure-7 it is clear that the proposed MQO approach is better than other approaches in terms of energy efficiency, time utilization and cost. The results and discussion section provides the justification on MQO.

7. RESULTS AND CONCLUSIONS

The total number of nodes deployed in the network is 200, 500, 1000 and 2000 for verifying the performance of the MQO approach. The three possible of the results are compared for various number of nodes mentioned above are the number of incoming queries, same type of queries and number of queries optimized. From this analysis, it shown in Figure-1, that the number of optimized queries are lesser than the total number of incoming queries. From this it clear or this clearly means that the energy and the time taken to process the optimized queries are much less. MQO has a great advantage that it can have communication only with the trusted nodes in the network. If the node is trusted node, then it start processing on the queries comes from those trusted nodes. From this it can be seen or hence MQO is also acting as a good anomaly detection and prevention system in WSN which eliminate the un-trusted nodes from communication. From the overall nodes requesting for communication, MQO chooses only best trusted nodes for communication, through this MQO can save more energy in terms for query processing and optimization and it is shown in Figure-2. It is well known that LEACH protocol selects only highest energy based nodes for data transmission. Similarly MQO also selects the best nodes [having highest energy] for query processing. Whenever data transmission happens among higher energy based nodes, it can increase the network life time. Figure-4 shows that the MQO filter out the nodes for query processing which are having more energy than the other nodes in the processing.

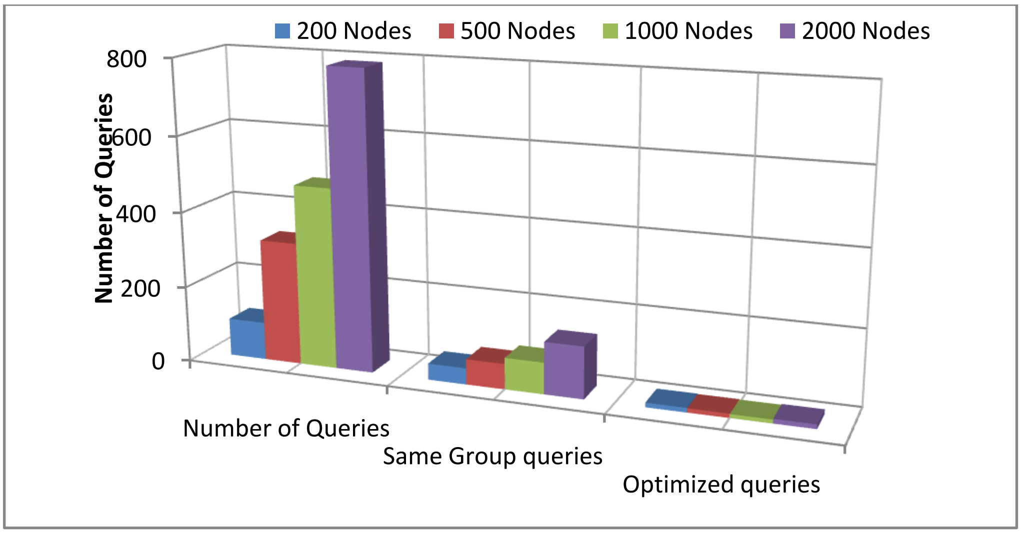

Figure 1: Number of Queries Generated vs. Number of Queries Optimized

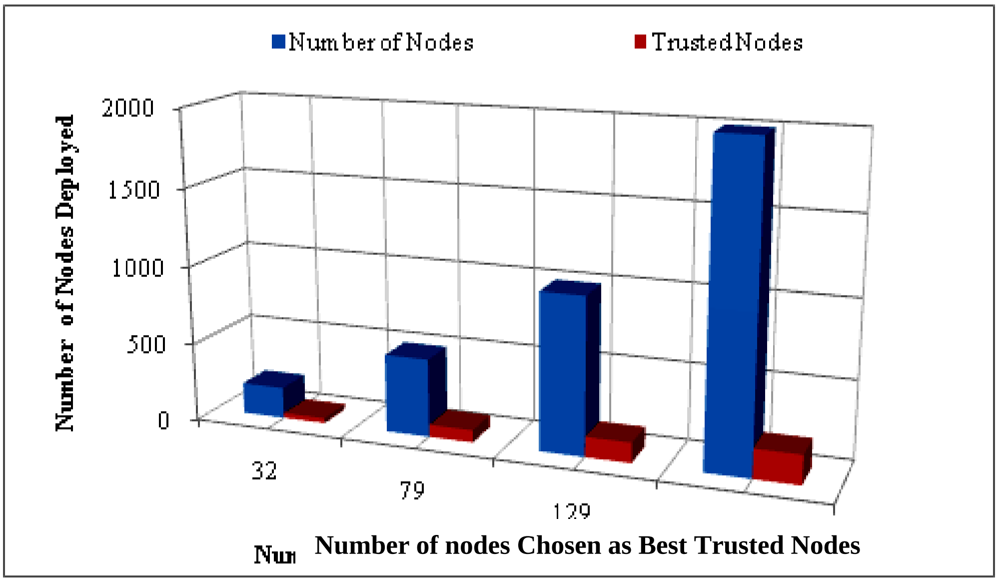

Figure 2: Number of Nodes in the Network vs. Number of Trusted Nodes Optimized

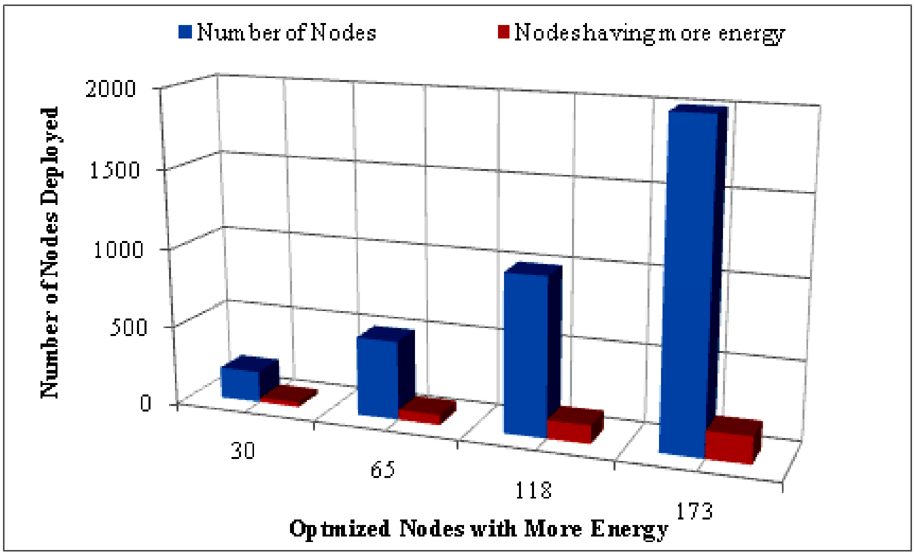

Figure 3: Number of Nodes in Network vs. Number of Optimized Nodes in terms of Highest Energy

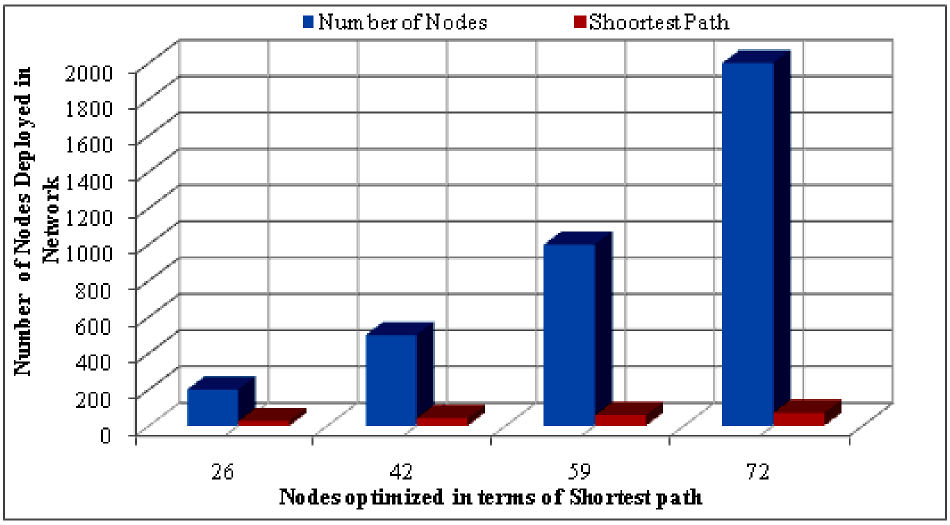

Figure 4: Number of Nodes in Network vs. Nodes Optimized in Terms of Shortest Distance

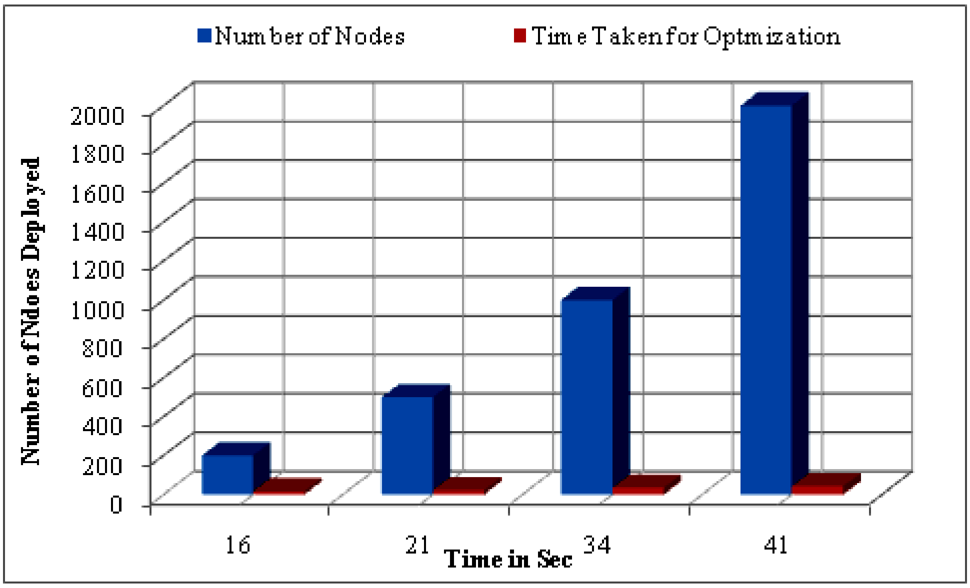

Figure 5: Number of Nodes Deployed in Network vs. Nodes Optimized in terms of Time

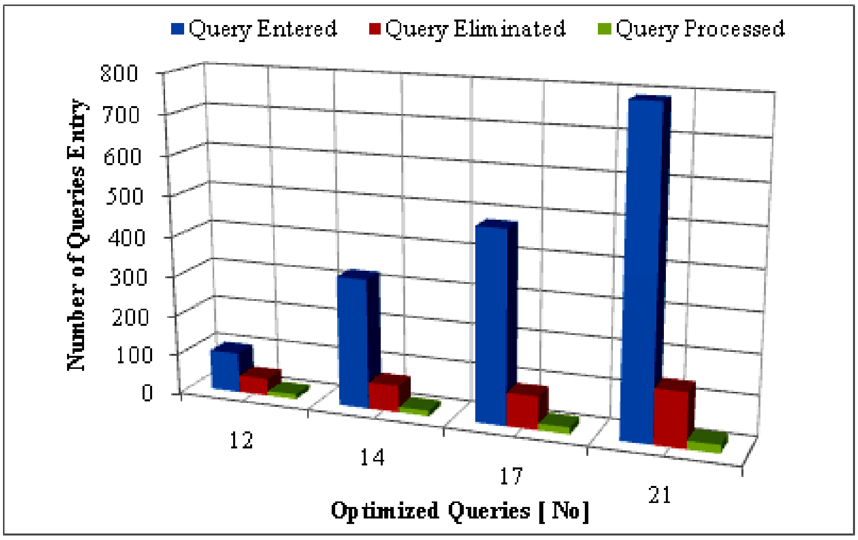

Figure 6: Number of Queries Entered vs. Number of Queries Optimized

In the same manner, MQO verifies the time taken for fetching and sending the data among the nodes, distance among the nodes and the BS, and availability of the data to reduce the amount of energy consumption. Figure-4 shows the nodes selected in terms of lesser distance, Figure-5 shows that lesser time taken for query process and Figure-6 shows that total number of queries raised in the network and number of queries are optimized for 100, 500, 1000 and 2000 nodes deployed in the network. From the above results it is established that MQO functions well and employs energy efficient based data transmission in wireless sensor network.

8. CONCLUSIONS

This paper addresses the issues in existing approaches and how the proposed MQO approach overcomes the issues such as time, cost and energy. For improving the energy efficient data gathering and data processing in WSN, this approach is based on shortest path, trustable node and less cost involving energy efficient query processing. Multiple aggregate queries take more time and memory to gather and aggregate. But the proposed MQO algorithm is efficient for query optimization and it has also been tested under various scenarios. Using these three strategies, MQO saves time and energy consumption, finally obtaining energy efficient data gathering and processing in WSN. Further this approach can be extended to distributed, controlled heterogeneous wireless sensor networks as well.

9. REFERENCES

[1] Divyakant Agrawal,Amr El Abbadi Lin Qiao, "Supporting Sliding Window Queries for Continuous Data Streams," in 15th International Conference on Scitific and Stastical Database management, 2003.

[2] J. J. P. C. Rodrigues L. M. L. Oliveira, "Wireless Sensor Networks: a Survey on Environmental Monitoring," Journal of Communications (JCM), vol. 6, no. 2, pp. pp. 143-151, April 2011, DOI: 10.4304/jcm.6.2.

[3] Amir Massoud Bidgoli, Mohammad Hossein Yektaie Arash Nikdel, "A new Scheduling Mechanism Inspired of Artificial Immune System Algorithm for Wireless Sensor Networks," International Journal of Smart Home, vol. 5, no. 4, October 2011.

[4] Johannes, Gehrke Yong Yao, "Cougar Approach to In-Network Query Processing in Sensor Networks," Sigmod Record, vol. 31, no. 3, September 2002.

[5] Joel J. P. C. Rodrigues,Mbaye Sene, Jaime Lloret Ousmane Diallo, "Distributed Database Management Techniques for Wireless Sensor Networks," IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 2013.

[6] Imane Aly Saroit, Hesham El-Mahdy, Eid Emary Marwa Sharawi, "Routing wireless sensor networks based on Soft computing paradigms: survey," International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI , vol. 2, no. 4, August 2013.

[7] Sharmila Devi Vaidehi V, "Distributed Database Management and Join of Multiple Data Streams in Wireless Sensor Network using Querying Techniques," in IEEE-International Conference on Recent Trends in Information Technology, 2011.

[8] Gondal, I., Kamruzzaman J. Rashid M, "Share-Frequent Sensor Patterns Mining from Wireless Sensor Network Data," IEEE Transactions on Parallel and Distributed Systems, vol. PP, no. 99, 2014.

[9] Giri C. Chowdhury S., "Data collection point based mobile data gathering scheme with relay hop constraint ," in Advances in Computing, Communications and Informatics (ICACCI) , Mysore, 2013, pp. 282 - 287.

[10] Jaime Lloret, Miguel Garcia and Jose F. Toledo S. Sendra, "Power saving and energy optimization techniques for Wireless Sensor Networks," Journal of Communications, vol. 6, no. 6, pp. Pp. 439-459, August 2011. doi:10.4304/jcm.6.6.439-459.

[11] Yong Yao, Alan Demers, Johannes Gehrke, Rajmohan Rajaraman Niki Trigoni, "Multi-query Optimization for Sensor Networks," in Distributed Computing in Sensor Systems.: Springer Berlin Heidelberg, 2005.

[12] L.D. Chagas and Univ. do Estado do Rio Grande do Norte, Lima E.P, Neto, P.F.R. Dept. de Inf., "Real-Time Databases Techniques in Wireless Sensor Networks ," in Networking and Services (ICNS), 2010, pp. 182 - 187.

[13] A. Lopes, D. Meira, R. Vasconcelos, and R. Menezes Brayner, "An adaptive in-network aggregation operator for query processing in wireless sensor networks," The Journal of Systems and Software , vol. 81, no. 3, pp. 328–342, 2008.

[14] Lars Kulik, EgemenTanin, Muhammad Umer, "Optimizing query processing using selectivity-awareness in Wireless Sensor Networks," Computers, Environment and Urban Systems, vol. 33, pp. 79–89, 2009.

[15] Feifei Li, George Kollios, John Byers Jeffrey Considine, "Approximate aggregation techniques for sensor databases," in 20th International Conference on Data Engineering, 2004, pp. 449-460.