

A Framework of Energy Efficient Query for Spatial Data in Wireless Sensor Networks

Zhanyang Zhang
Computer Science Department
College of Staten Island / City University of New York
2800 Victory Blvd., Staten Island, NY 10314
zhangz@mail.csi.cuny.edu

Abstract — *Due to many resource constraints of wireless sensor networks and characteristics of sensor data stream, some of the data management techniques that work well for traditional database are insufficient for managing and processing sensor network data. In this paper we present a cluster based query protocol for wireless sensor networks which uses self-organized sensor clusters to register queries, process queries and disseminate data within the sensor network. This protocol can provide an integrated solution to address some of the challenge problems of wireless sensor networks, including locating sensors, energy efficient data query processing, and fault tolerant network operations.*

Keywords: Wireless Sensor Networks, Sensor Clusters, Cluster Based Query, Spatial Data, Real-Time Data Stream.

1. Introduction

Wireless sensor networks extend our capability to explore, monitor and control the physical world. Each sensor monitors and collects data about the world at large. A wireless sensor has limited data processing, storage and communication capabilities, most critically the limitation of energy resources. Due to the short battery life that powers individual sensors, the lifetime of sensor networks is severely restricted [1, 2]. In addition to this limitation, sensor data behaves differently in many ways from the data in traditional database. Each sensor generates a real-time data stream about its environment. The data volume can overload the limited storage and communication capacity of a sensor network. It is unpractical to store a large volume of raw data locally at the data sources or to transmit the data over the sensor network to a central depository [3, 4].

In many environment monitoring applications, sensor data is highly spatial in its nature. Thus it requires correlating the data with the location of the data source in order to interpret the data in a meaningful way. On

the other hand, sensors could be randomly deployed in large quantity and they might not aware of their own locations within the deployed region [5, 6]. Last but not least, sensor failures and disconnections are part of reality in many wireless sensor network applications.

All of these present a set of challenging problems for wireless sensor network design and operation. Many of the data management techniques that work well for traditional database are insufficient for managing and processing sensor network data. Motivated by these challenges, we focus our research on one of the major types of wireless sensor network application, environment monitoring and control in open spaces, where sensors are randomly deployed in high density and an individual sensor has no knowledge of its location. This type networks require a long duration of reliable operations with consideration of sensor failures and intermittent connections. One of the application examples could be pollution monitoring and surveillance systems that collect real-time air and water quality data to satisfy state government air-water pollution monitoring requirements.

Our research objective is to develop an integrated solution to address the problems of locating sensors, energy efficient data query processing, and fault tolerant network operations. In the proposed protocol there are four components working together to address these problems, namely, a Stimulating Signal Model (SSM) that uses a laser beam as a stimulating signal to trigger sensors with in one hop of communication distance to form clusters; a Sensor Cluster Location Algorithm (SCLA) that estimates sensor cluster locations and elects a cluster head; a Sensor Cluster Maintenance Algorithm (SCMA) that can sustain sensor cluster function by replacing failed sensors due to random failures or running out of power; and a mechanism to register, process queries, and disseminate data over wireless sensor networks.

The rest of the paper is organized as follows. In section 2, we introduce the SSM model as an operational model for the protocol to form sensor clusters. The SCLA algorithm and its scalability analysis are given in section 3. We propose the SCMA

algorithm in section 4. We present our data query protocol and outline the mechanisms of query registration and query processing for energy efficiency and network data traffic reduction in section 5. We show our scalability and performance analysis and simulation study results in section 6. Finally, we conclude this paper with discussions about our current and future research works in section 7.

2. The Stimulating Signal Model

In the SSM model, the sensor network is deployed and operated under the conditions defined by the following set of assumptions.

1. Sensors – Each sensor has sensing, data storage, data processing and wireless communication capability that is equivalent to that of a MICA Mote (UC Berkeley) [7, 8]. Each sensor covers a communication cell and a sensing cell defined by radius R_c and R_s respectively. All of the sensors can sense optical signals (delivered by laser beams).
2. Sensor network – A sensor network consists of a set of homogeneous sensors. These sensors are static and not mobile. Sensors can communicate with each other via wireless channels in single or multiple hops, thus they form an ad-hoc network. There are one or more base stations located outside the sensor region but near the boarder of the sensor network with wired or long range wireless communication links to the Internet for collecting data or disseminating queries and control commands to the sensor network.
3. Deployed region – Sensors are deployed over an open-space area where there is an uninterrupted line-of-sight path for laser beams to reach ground level. A virtual grid marks this area. Each cell in the grid is a D -by- D square. To simplify the model, we assume there are no sensors located on the borders of cells.
4. There is a lightweight location guided laser designator system that can project a laser beam to a given location (x, y) with acceptable accuracy [9].
5. To ensure the coverage and connectivity of a sensor network, the model requires that D , R_c and R_s satisfy the following conditions:
 - To ensure a sensor anywhere in a cell can cover the cell, it must satisfy the condition, $R_s^2 \geq 2D^2$. In our model we assume $R_s^2 = 2D^2$.
 - To ensure a sensor anywhere in a cell can communicate with a sensor anywhere in its neighboring cell, it must satisfy the condition, $R_c \geq 2R_s$. In our model, we assume $R_c = 2R_s$.

Fig. 1 below shows the parameters that define virtual grids with 4 neighboring cells.

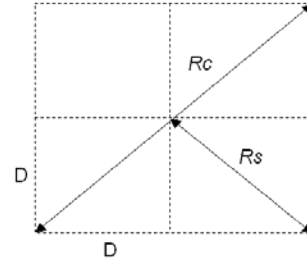


Fig. 1 Parameters that define virtual grids

A typical sensor network operates in five phases: the planning phase, deployment phase, post-deployment phase, operation phase and post-operation phase. In the planning phase, a site survey is conducted to evaluate deployment environment conditions and then to select a suitable deployment mechanism. In the deployment phase, sensors are randomly deployed over a target region. In the post-deployment phase, the sensor network operators need to identify or estimate the location of sensors. The operation phase involves the normal operation of monitoring tasks where sensors generate data and communication the result to the base station via multi-hops wireless links. The post-operation phase involves shutting down and preserving the sensors (setting the sensors to sleep mode) for future operations or destroying the sensor network.

Our model works during the post-deployment phase to identify sensor clusters (sensors that are located in the same cell of the virtual grid) and to estimate sensor cluster locations [10]. It can verify the sensor network coverage using the sensor cluster location data and the number of members in the cluster [11]. Here we describe how the model works:

Every point in the region can be represented by a pair of (x, y) coordinator values. A sensor has three possible states, U (unknown), H (cluster head) and M (cluster member). Initially all sensor states are set to U. During the post-deployment phase, an object flies over the deployed region and projects a laser beam to the center of a grid cell (X_c, Y_c) . The sensors nearby will sense the signal. The sensor readings are stronger if they are closer to the projected laser beam. The sensor with the strongest reading is identified as the cluster head (state=H). All sensors that have a reading greater than λ (λ -cut) and are one hop away from the cluster head become members of the cluster (state=M). The ideal value for λ should maximize the possibility of including a sensor in the cluster if it is in the cell, and minimize the possibility of including a sensor in the cluster if it is outside the cell. An optimal value of λ can be obtained through experiments and simulations. We believe it is a reasonable assumption that the light signal decay model should be similar to

the attenuation of radio waves between antenna and wireless nodes close to the ground. Radio engineers typically use a model that attenuates the power of a signal as $1/r^2$ at short distances (r is the distance between the nodes), and as $1/r^4$ at longer distances [12]. In our model the size of the virtual grid is small and we can assume the light signal attenuation model as $1/r^2$.

Fig.2 shows the cluster formed in grid cell 5. The black dot indicates the sensor node is a cluster head in the cluster.

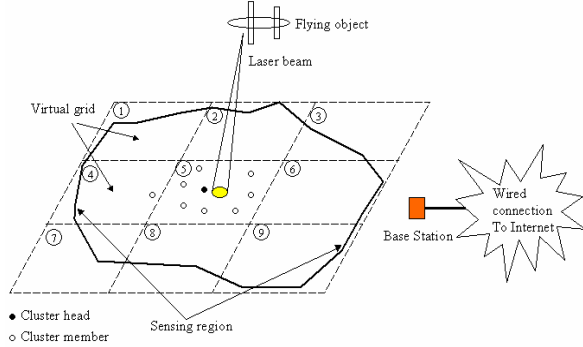


Fig. 2 SSM model for locating sensor clusters

In this paper, we assume that a laser beam is projected to one cell at a time with a cluster forming time interval, T , for each cell. T should be just long enough to allow the sensors in a cell to form a cluster, but not so long as to cause unnecessary delay for the operations between cells. In general, T is a function of sensor density - n (the number of sensor nodes in a cell), radio propagation delay τ and MAC layer back off delay β with p as a probability of package collision in the form of:

$$T = n^2 \beta p + n \tau + k \quad (1)$$

Where k is a constant

With this assumption, a sensor node can only belong to one cluster, since once it joins a cluster it will not respond to the stimulating signals projected to other cells.

3. The Sensor Cluster Location Algorithm (SCLA)

Based on the model described above, we present the formal algorithm with the following steps:

1. Let t_0 be the time when the laser beam is projected to a cell and let T be the time interval for cluster formation. Let $t = t_0$ when starting the algorithm.
2. While $t > t_0$ and $t < (t_0 + T)$ repeat step 3 and 4.

3. For each sensor with unknown status that has detected the signal, if the sensor reading is greater than λ , then it will broadcast a message with the sensor id and sensor reading (SID, Value) to its neighbors within one hop of communication. Otherwise keep silent.
4. When receiving a message, a sensor with unknown status acts according the following rules: Rule1 - If the reading value of the received message is greater than its own reading and its own reading is greater than λ , then it will set the state=M (a cluster member) and reset its local memory. Rule2 - If the reading value of received message is less than its own reading and its own reading is greater than λ , then save the message (SID, Value) in its local memory.
5. For a sensor that still has unknown status, if its own reading is greater than λ it will set the state=H (a cluster head).
6. The cluster head sends the cluster membership information (SID, Value) pairs, which were saved at step 4, to the closest base station.
7. Project laser beam to the next cell and repeat the above steps 1 to 6 until all the cells are visited.

Our analysis shows the SCLA algorithm performs and scales well when sensor network size increases in terms of both the number of cells in a grid and the total number of sensors in a cell. Let L be the largest number of communication hops from a cluster head to the closest base station. Let M be the total number of cells in the grid. Let N be the total number of sensors deployed. Let $n(i)$ be the number of sensors in cell(i). The cost of SCLA algorithm in terms of the number of messages transmitted is given as:

$$Cost(L, M, N) \leq \sum_{i=1}^M n(i) + M * L \quad (2)$$

$$\text{where } \sum_{i=1}^M n(i) = N$$

If we assume sensors are uniformly distributed, then we have:

$$Cost(L, M, N) \leq M*(N/M) + M*L = N + M*L \quad (3)$$

It is equivalent to $O(N)$ when M and L are significantly smaller than N , which is true in most high-density sensor networks.

These results show that it is feasible to deploy large amount redundant sensor nodes in order to compensate for high rates of sensor failure with limited overhead cost. The overhead cost in terms of messages is a linear function of total sensor nodes.

4. The Sensor Cluster Maintenance Algorithm (SCMA)

The SCMA algorithm consists of three components, namely, cluster member component, cluster head component, and base station component. These components work together to achieve these functions:

1. Coordinating a subset of cluster members as active nodes to perform required tasks while setting the rest of cluster members as sleep nodes to conserve energy.
2. To replace the nodes that failed unexpectedly or whose energy level falls below a certain threshold to guarantee required coverage.
3. It should warn the network operators if the required coverage is going to be compromised.

After the cluster is formed according to the SCLA algorithm, all cluster members are in active mode ($m=A$) by default. A member listens to the initialization message, (I, sensorlist), from the cluster head. A member turns into sleep mode ($m=S$) unless its SID is on the sensor list in the initialization message. We use finite state automata to represent the behavior of each component. Particularly we adopted a special type of finite state automata, called Mealey machine, which is formally defined by Hopcroft and Ullman [13] as below:

Let a finite state automata, $F = (Q, \Sigma, \Delta, \delta, \gamma, q_0)$ where:

- q_0 is the initial state ($q_0 \in Q$),
- Σ is a set of inputs,
- Q is a set of states,
- Δ is a set of outputs,
- δ is a state transition function: $Q \times \Sigma \rightarrow Q$,
- γ is an output function: $Q \times \Sigma \rightarrow \Delta$.

In the rest of the section, we will use state transition diagrams to describe each SCMA algorithm components. To help understanding and tracking the state transition diagrams, we list all the messages and their definitions in the table below.

Tab. 1 Definition of messages for SCMA algorithm

Message	From	To	Description
(A, SID)	Head	Member	Activation- Activate a member (SID)
(Ack, SID)	Member	Head	Acknowledgement- A member (SID) is active
(Ack, SID)	Member	Base	Acknowledgement- a member (SID) is a new head
(E, HID)	Head	Base	Emergency message to the base station
(H, HID)	Head	All members	Broadcast "hello" message

(H, SID)	Member	Head	Reply to "hello" message by a member (SID)
(I, Sensorlist)	Head	All members	Broadcast initial active sensor list
(N, SID)	Base	Member	Appoint a member (SID) to be the new head
(R, SID)	Member	Base	A member requests the cluster status
(Status, HID)	Head	Base	Send cluster status to the base station
(Status, SID)	Base	Member	Send cluster status to a member (SID)
(W, HID)	Head	All members	Broadcast "wakeup" message
(W, SID)	Member	Head	Reply to "wakeup" message by a member (SID)
Tx / e	Time trigger – while in state x for time interval Tx then return to previous state without output. $x \in (c, r, s, w, p, u, n, g)$.		
Pa / e	Event trigger – power level below a.		

4.1. The Cluster Member Component

We use two state transition diagrams, one represents a cluster member in active mode and another one represents cluster member in sleep mode. Fig. 3 defines the behavior of an active member node within the SCMA algorithm and how it can change to sleep mode or become a cluster head when a cluster head has failed.

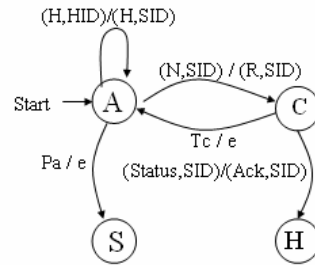


Fig. 3 State transition diagram for active members

Fig. 4 describes the behavior of a sleep member node and how it can change to active mode.

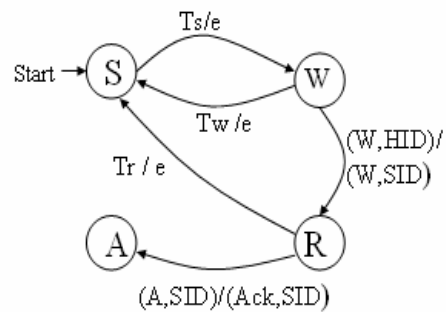


Fig. 4 State transition diagram for sleep members

4.2. The Cluster Head Component

The cluster head component coordinates members between active and sleep modes. It updates cluster status and synchronizes the cluster status with the base station. It uses an active node counter (Ac) and a sleep node counter (As) to track active members and sleep members. The Fig. 5 below describes how a cluster head works. If the cluster head issues an emergency message, (E , HID) to base station if the required coverage is going to be compromised.

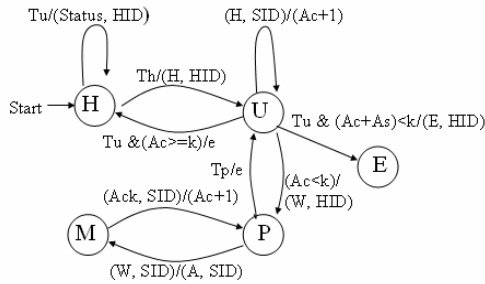


Fig. 5 State transition diagram for cluster head

4.3. The Base Station Component

In the SCMA algorithm, the function of a base station component is to oversee the status of clusters and the condition of networks. Fig. 6 shows how it works.

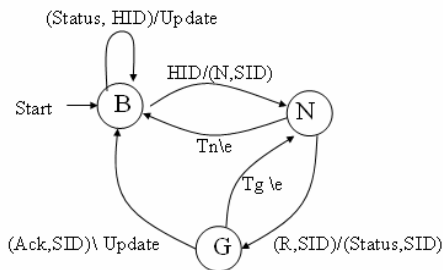


Fig. 6 State transition diagram for base station

5. The Cluster Based Query Protocol

We proposed a cluster-based protocol (CBP) for energy-efficient data management and query processing, which works particularly well for the queries of spatial data in sensor networks. CBP uses cluster heads as data storage and aggregation points. Instead of sending large amounts of raw data over a network to reply to a query, each cluster head collects and filters data from its member sensors. When data queries are entered into sensor networks from the base station, CBP can bring queries to the data source by disseminating and registering queries at the cluster heads selected by the spatial qualifiers defined in the queries. This is achieved using the cluster location information obtained from the SCLA algorithm. Many published research works have shown that transmission of data consumes the most energy in wireless sensor

networks [14, 15]. Therefore, data communication between sensor nodes is a key measurement for energy efficiency. With CBP protocol, energy efficiency is achieved by reducing data transmissions over the network during the course of the data collection and query processing. For example, we can enter a SQL-like query into a sensor network to detect fire hazard conditions in certain high risk areas of a forest as:

```
Select temperature, humidity
From north-west corner
Where temperature > T and humidity < H
Duration 48 hours
Period every hour
Starts 22:00.00 pm June. 8th, 2006
```

First, queries are routed and registered at the targeted cluster heads without flooding the network. Second, data are aggregated from the member sensors in a cluster to the cluster head. The cluster head computes average value of the required data and only sends over the query result if it meets the query thresholds (Fig. 7).

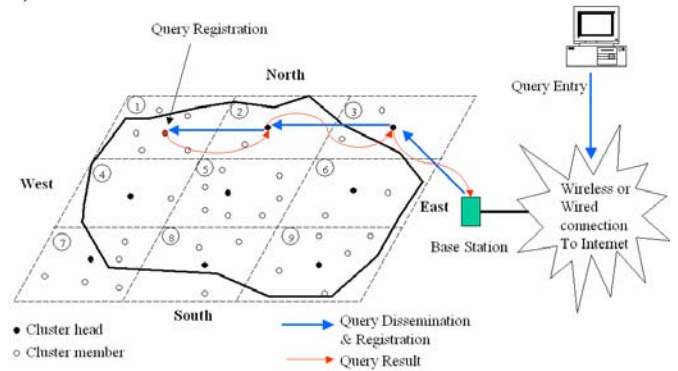


Fig. 7 Cluster Based Protocol for Spatial Data

CBP increases network reliability and enhances operation robustness due to the fact that there are multiple sensors in each cluster with redundant and overlapping sensing and communication capabilities. In case of one or more sensor nodes failed or disconnected communication links, there are other sensors and links within the cluster that can support the routine sensor network operations.

6. Simulation Study and Analysis

This research work is still in progress. Our analysis and simulation results show that the SSM and SCLA algorithm scale well to large numbers of sensors as well as large monitoring areas. The CBP protocol can achieve greater energy savings as the scale of the network increases in terms of the number of sensors, the size of the monitoring areas and the length of the operations.

In order to validate the model and algorithm presented in this paper and to gain insights into how

the algorithm works, we conducted simulation studies using NS2 simulator with Monarch Extensions to NS2 [16,17].

Our simulations are implemented in two scenarios. The first scenario involves simulating a single grid cell with different sensor densities (number of sensors in the cell). The focus of this simulation is to study the performance and scalability of our model against sensor density. In the second scenario, we take the same measurements from a multi-cell grid simulation with considerations of both sensor density and the size of the deployment area in terms of the number of cells. The purpose of this simulation study is to understand the performance and scalability of our model in a multi-cell grid.

We set the cell dimension to 10 meters for all the simulations presented in this paper. Our simulation tests indicate that the outcomes are not as sensitive to the cell dimension as they are to sensor density. We let the number of sensor nodes be varied from 10 to 80 in increments of 10. In our simulation model, we set the propagation delay between two nodes as 10 ms ($\tau = 10\text{ms}$). We use multicast in UDP protocol to simulate sensor node broadcast in one hop distance. We set p as the probability for a node to receive the broadcast message successfully (p in the range 0 to 1). In our simulation we set $p=0.99$. The message package size is set to 128 bytes and the bandwidth between two nodes is set to 2mbs. To simulate IEEE 802.11 MAC layer data collision protocol, we introduce a back-off time delay, a random number between 50 and 100ms, which is assigned to a node when it detects that a channel is busy. The node will back-off for a delay interval before it tries to broadcast again.

The simulation results capture the key measurement, the number of messages being transmitted during the cluster formation at the cell. All of the simulation results represented below are the average of five simulation runs.

Fig. 8 shows the number of messages being transmitted in a single cell grid. It compares the analytical result with simulation results. It indicates that the cost of message transmissions is close to a linear function of n , the number of sensor nodes in the cell.

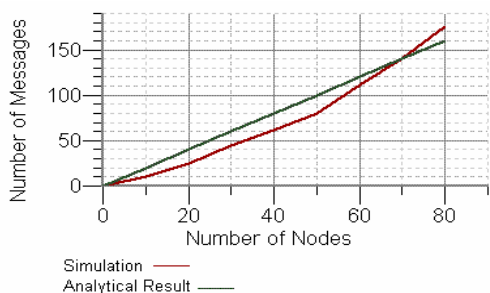


Fig. 8 Messages transmitted in a single cell

To better understand cluster membership distributions and study the impact of λ values on member selections, we used a single grid cell of 10 meters by 10 meter with λ values in the range [0.02, 0.04]. The simulation result in Fig. 9 shows the percentage of sensors that are dropped from the cluster for the cell as the value of λ changes. It shows the higher λ value leads to more sensor nodes being excluded from the cluster.

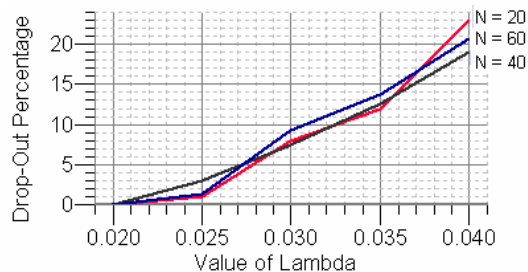


Fig. 9 Membership distribution over γ values

The simulation results below are for a multi-cell grid scenario with the same key measurement as we presented for a single cell grid. Fig. 10 shows that with a fix number of sensor nodes, the number of messages being transmitted actually drops as the number of cells increase. This is not exactly what we expected as the cost function defined in (3). Because there are fewer collisions as sensor density decreases. This simulation result indicates that the number of messages being sent is more sensitive to the density in each cell than the number of cells in a grid.

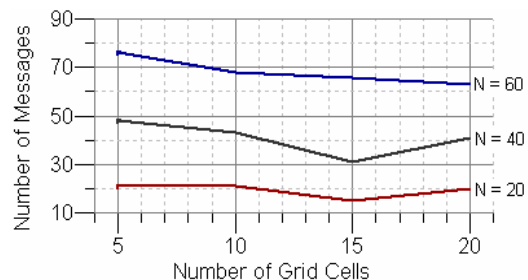


Fig. 10 Performance and scalability simulation in a multi-cell grid measured in messages transmitted

Fig. 11 presents an interesting measurement, the percentage of sensors wrongly claimed by clusters. It is closely correlated with the γ values. The ratio of disputed sensors drops or stays at the same level after the number cells reach 16 due to the decrease in sensor density.

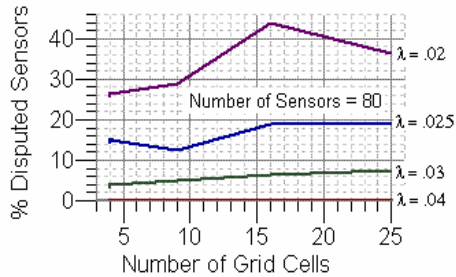


Fig. 11 Number of disputed sensors with γ values

Fig. 12 shows the percentage of sensors that are unclaimed in a multi-cell grid with 80 sensors total.

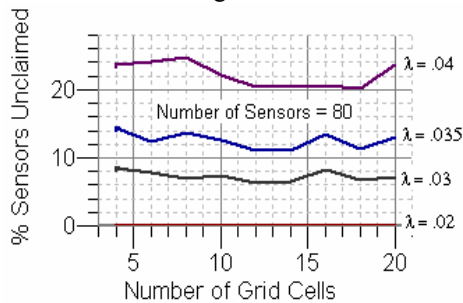


Fig.12 Percentage of unclaimed sensors

6. Conclusion

In this paper we present a cluster based query protocol for wireless sensor networks that provides a solution to address the sensor location, energy-efficiency and fault-tolerant issues. This research work is still in progress. We are aware of the fact that cluster heads consume more energy than other member sensors due to extra duties. We proposed a distributed Sensor Cluster Maintenance Algorithm (SCMA) to address these issues. SCMA can maintain the required cluster membership by selecting a subset of sensor nodes to be active and perform sensing tasks, while the rest of the sensor nodes are in sleep mode. A subset of sleeping sensors can be activated to replace the role of one or more active sensor nodes, including the cluster head, in case of node failures or energy level failing below certain threshold. This will greatly enhance sensor network reliability and prolong network operation lifetime. Obviously, the stimulating signal scheme employed in the model requires an uninterrupted line-of-sight path for laser beams. However for many environmental monitoring applications, this is not a major problem.

We are developing and testing a query optimization algorithm to aggregate queries that are targeted to the same clusters with overlap query results. This will enable several queries to merge into one query that is disseminated, registered and processed once to further reduce data traffics over sensor networks.

7. Acknowledgement

This work was supported (in part) by a grant from The City University of New York PSC-CUNY Research Award Program (grant#:68358-0037), a grant from CUNY Research Foundation (80210-14-04) and the funding from the New York State Dormitory Authority (GS7022).

8. References

- [1] I.F. Akyildiz et al. "A Survey on Sensor Networks". IEEE Communications Magazine, August 2002, pp. 102-114.
- [2] C. Jone, et al. "A Survey of Energy Efficient Network Protocols for Wireless networks", Journal of Wireless Networks, Kluwer Academic Publishers, 2001, Vol. 7, pp. 343-358.
- [3] C. Intanagonwiwat, et al. "Directed diffusion: a scalable and robust communication paradigm for sensor networks", Mobile Communication and Networking, 2000, pp. 56-67.
- [4] S. Kim, et al. "Data Dissemination over Wireless Sensor Networks", IEEE Communication Letters, 2004.
- [5] A. Savvides et al. "Dynamic Fine-Grained Localization in Ad-hoc Networks of Sensors", ACM SIGMOBILE, Rome, Italy. July 2001.
- [6] J. Hightower and G. Borriello. "Location Systems for Ubiquitous computing", IEEE Computer, August 2002, pp. 57-66.
- [7] "MICA Wireless Measurement System", <http://www.xbow.com/Products/>
- [8] "University of Berkeley Mote", <http://webs.cs.berkeley.edu/>
- [9] "Lightweight Laser Designator Rangefinder", <http://www.defensedaily.com/progprof/army/ldr.pdf>
- [10] Z. Zhang and W. Gavin, "Using Stimulation Signal to Locate Sensor Clusters", the 11th European Wireless Conference, Cyprus, April, 2005, pp. 719-724.
- [11] Z. Zhang and W. Gavin, "A Cluster Based Feedback Model for Sensor Deployment Management and Coverage Verification in Wireless Sensor Network Applications", the 2005 annual International conference on Wireless Networks, Las Vegas, June, 2005, pp. 441-447.
- [12] J. Broch, et al. "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols". Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking, October, 1998, Dallas, Texas.
- [13] Hopcroft J. and Ullman J., "Introduction to Automata Theory, Language, and Computation". Addison Wesley publishing company, 1979. Page 42 – 44.
- [14] M. Cardei et al.. "Wireless Sensor Networks with Energy Efficient Organization", Journal of Interconnection Networks, Vol 3, No 3-4, 2002, pp. 213-229.
- [15] V. Shnayder et al. "Simulating the Power Consumption of Large-Scale Sensor Network Applications", ACM SenSys '04, Baltimore Maryland, November, 2004.
- [16] "The Network Simulator ns-2: Documentation", <http://www.isi.edu/nsnam/ns/ns-documentation.html>.
- [17] "Monarch Extensions to ns", <http://www.monarch.cs.cmu.edu/>