Fundamental Scaling Laws for Energy-Efficient Storage and Querying in Wireless Sensor Networks

Joon Ahn and Bhaskar Krishnamachari
Department of Electrical Engineering
University of Southern California
Los Angeles, California, USA
joonahn@usc.edu, bkrishna@usc.edu

ABSTRACT

We use a constrained optimization framework to derive fundamental scaling laws for both unstructured sensor networks (which use blind sequential search for querying) and structured sensor networks (which use efficient hash-based querying). We find that the scalability of a sensor network's performance depends upon whether or not the increase in energy and storage resources with more nodes is outweighed by the concomitant application-specific increase in event and query loads. Let m be the number of events sensed by a network over a finite period of deployment, q the number of queries for each event, and N the size of the network. Our key finding is that $q^{1/2} \cdot m$ must be $O(N^{1/4})$ for unstructured networks, and $q^{2/3} \cdot m$ must be $O(N^{1/2})$ for structured networks, to ensure scalable network performance. These conditions determine (i) whether or not the energy requirement per node grows without bound with the network size for a fixed-duration deployment, (ii) whether or not there exists a maximum network size that can be operated for a specified duration on a fixed energy budget, and (iii) whether the network lifetime increases or decreases with the size of the network for a fixed energy budget. We discuss the practical implications of these results for the design of hierarchical two-tier wireless sensor networks.

Categories and Subject Descriptors: C.2.2 Computer Communication Networks: Network Protocols

General Terms: Design, Performance, Theory

Keywords: Modeling, Wireless Sensor Networks, Energy Efficiency, Theory, Performance Analysis, Querying, Scalability

1. INTRODUCTION

Wireless sensor networks are envisioned to consist of large numbers of embedded devices that are each capable of sensing, communicating, and computing. While the network as a whole is required to provide fine resolution monitoring for an extended period of time, the individual embedded devices face some fundamental constraints. They are typically deployed with limited battery supplies and, because of their form factor and low cost, may also have limited data storage capability. The goal of this paper is to understand the conditions under which a query-based data-centric sensor network [1] can be operated in a scalable manner despite these constraints on energy and storage.

We consider both unstructured and structured varieties of data-centric querying along with replicated storage in this paper. In unstructured querying schemes, the node issuing the query does not know in advance where any copy of the requested event information can be found. The query dissemination is therefore a form of blind search (this can take the form of an expanding ring search or a sequential trajectory search). In structured querying schemes, a hash or index is used so that the querying node knows exactly where the nearest copy of the requested event information can be found. In such networks, there is a trade-off between the energy costs of replicated storage and querying that is determined by the number of replicas created for each event. A large number of replicas results in lowered query cost at the expense of greater storage cost, and vice versa. We can formulate an optimization problem whose aim is to select the optimum number of replicas that minimizes the total energy cost of querying and storage, subject to constraints on storage. We use this optimization problem as a tool to identify the conditions, in terms of the numbers of events and queries, under which query resolution can be performed in a scalable manner despite constraints on storage and energy.

We find that operating a network in a scalable fashion essentially requires that the traffic load due to additional events and queries be outweighed by the improvement in energy and storage resources obtained as the network size increases. Note that the scaling of event and query activity with network size is application specific — e.g., in many applications there may be only a constant number of queriers regardless of the network size, but the number of events detected grows linearly with the covered area; in other applications, the number of querying nodes may increase in some fashion with the network size, while the events detected remain constant.

The following are the key contributions of this work:

- We present models for the search and replication costs for structured and unstructured networks for two-dimensional grid and random network deployments (see section 3); then formulate and solve an optimization problem to determine the optimal number of replicas in each scenario to minimize the total energy cost subject to storage constraints (see section 4).

- We derive the event-query scaling conditions to ensure that the required storage per node does not grow without bound as the network size increases (see section 5). Let N be the size of the sensor network, m the total number of events that are generated in the network during its fixed period of opera-

MobiHoc'06, May 22–25, 2006, Florence, Italy.

tion, and q the number of queries per event. We find that for unstructured querying, $q^{1/2} \cdot m$ must be $O(N^{3/4})$ to avoid requiring unbounded storage per node for efficient operation, while the equivalent condition for structured querying is that $q^{2/3} \cdot m$ must be $O(N)$.

- We derive the conditions under which the energy requirement per node remains bounded as the network size increases (see section 6). For bounded energy, we find that $q^{1/2} \cdot m$ must be $O(N^{1/4})$ for unstructured querying and $q^{2/3} \cdot m$ must be $O(N^{1/2})$ for structured querying. Note that the conditions for achieving the bounded energy are stricter than the conditions for achieving bounded storage. With regard to scalability, this suggests that energy constraints are fundamentally more limiting than storage constraints. Further, the conditions are so strict that even reasonable models for the scaling of event generation (e.g., having the number of events increase proportionally with the area covered) cannot be sustained by arbitrarily large networks.

- We investigate the scaling of network size when we have a fixed per-node energy budget (see section 7). We find that when the event-query rates scale faster than the above-stated conditions for bounded energy, there exists a maximum network size beyond which not all queries for events can be resolved within the period of deployment before the available energy is depleted. A finer-grained analysis reveals when the maximum network size increases as a concave function of the average per-node energy, and when it increases as a convex function. This is useful from a design perspective as it indicates whether investing in an increased per-node energy allocation results in super-linear or diminishing returns with respect to network size.

- We consider variable-time deployments on a fixed energy budget and examine how the network lifetime varies with network size (see section 8). We find that depending on the query-event scaling behavior, the lifetime can increase, remain constant or decrease with the addition of nodes to the network.

- We argue that limiting the network size to a maximum value can be interpreted as decomposing a larger network hierarchically into many multi-hop clusters of size smaller than this maximum value, such that queries are limited to events sensed and stored within the cluster. If the application should require that queries from farther off be resolved, then it is essential to create a wired second-tier which can transport queries across clusters with minimal energy overhead.

1.1 Related work

Our focus on scalability issues studied using order notation is certainly inspired by the well-known work on transport capacity of wireless networks by Gupta and Kumar [7], though we do not focus on wireless bandwidth limitations. There has also been some work on the asymptotic energy-constrained capacity of wireless sensor networks [8]. And some prior studies have looked at maximizing the lifetime of continuous data-gathering [9, 10, 11]. However, these studies are different in scope from our work which is focused on the scalability of wireless sensor networks that employ data-centric storage and querying.

There have been several interesting prior studies on analytical modeling of query strategies [12, 13, 14, 15]. The energy costs of data centric storage are compared with the two extremes of external storage and local storage in [12]. A hybrid push-pull query processing strategy is proposed and analyzed in [13]. Shakkotai [14] has presented a comparison of the asymptotic performance of three random walk-based query strategies, showing that a rendezvous-based sticky search has the best success probability over time. The optimal parameter setting for the comb-needles approach is analyzed in [15]. An analytical comparison of the comb-needles approach and data centric storage is provided in [16]. These studies have not developed fundamental scaling laws for data-centric querying with replicated storage with respect to the scaling of event and query loads — to our knowledge this is the first work on the topic.

2. ASSUMPTIONS

The following are the key assumptions in our work:

- N nodes are deployed with constant density in a two- dimensional square area. The constant density implies that if the network size is increased, the deployment area grows proportionally.

- Our results are applicable to both square grid and random deployments of nodes, because we show that they both have the same scaling of querying and storage costs except for different constants.

- The radio radius of a node is R for all nodes.

- The sensor network is deployed for a fixed application-specific time duration T.

- During this time duration, there are m atomic events that are sensed in the environment. The distribution of events is assumed to be uniform in the deployment area.

- A total of r_i copies of each event are maintained with a uniform distribution in the network by creating $r_i - 1$ additional replicas when the event is first sensed.

- For each event i, there are a total of q_i queries that are generated uniformly by the nodes in the network. Each query is a one-shot query (i.e. requires a single response, not a continuous stream), and is satisfied by locating a single copy of the corresponding event.

- We assume that the links over which transmissions take place are lossless (e.g., using blacklisting) and present no interference due to concurrent transmissions (e.g., due to low traffic conditions or due to the use of a scheduled MAC protocol).

- The total energy cost for storage and querying is assumed to be proportional to the total number of transmissions. This is reasonable particularly for sleep-cycled sensor networks where radio idle times are kept to a minimum.

- We assume that the storage at each node is a constant amount s, so that the total storage $S = s \cdot N$, where each event copy requires a unit of storage.

3. MODELING QUERYING AND REPLICATION COSTS

We now turn to developing mathematical models to quantify the cost of replication and search. We consider two types of data-centric querying techniques: structured and unstructured. In structured environments, the data is stored in the network and retrieved from it using a hash. This approach is exemplified by the geographic hash-table technique [2]. Thus in structured querying, the querying node is aware of the location of the nearest copy of the replicated event information and sends the query directly to this point to get a response. In unstructured environments, by contrast, there is no predetermined location where the querying node can send a query. Hence the query must be disseminated through a form of blind search. If latency is not a concern, efficient unstructured querying strategies involve expanding ring searches or sequential trajectories [3, 4].

It turns out that whether the network is deployed in an area uniformly with a random distribution of nodes or as a regular grid, the expressions resulting energy costs for storage and querying are the same, except for differences in coefficients. We present these coefficients for the two deployments in Table 1. Detailed derivations are presented in our technical report [6]. We present below instead some approximate first-order modeling with intuitive explanations for how these costs vary as a function of the network size N and the number of copies r, for a given event.

First consider the replication costs. In both the structured and unstructured case these are same. The average number of hops from random event locations in the network to random locations is proportional to $\sqrt{N}$ (since the N nodes are placed in a square area). Thus the cost of creating and placing $r-1$ replicas at random locations in the network from random event locations is:

$$C_{replication} = c_1 \cdot \sqrt{N} \cdot (r-1) \tag{1}$$

Let us then consider the search cost for a structured environment. If the number of copies is kept fixed, since the replicas are placed uniformly in the network, the distance (in hops) between the querying node to the nearest replica increases with the network size as proportional to $\sqrt{N}$. If, on the other hand, the network size is kept fixed, then as the number of replicas increases and continues to be placed in the two-dimensional area with a uniform distribution among the N nodes, the expected one-dimensional distance to the nearest replica decreases inversely proportional to $\sqrt{r}$. Thus we have the following:

$$C_{search,structured} = c_2 \cdot \frac{\sqrt{N}}{\sqrt{r}} \tag{2}$$

Finally, let us consider the search cost for an unstructured environment. The search is analogous to looking sequentially for the first of r specific objects of a desired type from a randomly ordered set of N total objects. It can be shown that the expected number of steps till the first object of the desired type is observed is given as :

$$C_{search,unstructured} = c_3 \cdot \frac{N}{r+1} \tag{3}$$

We have derived the above expression in previous work, for both random and grid settings [5, 16].

We note that in calculating the search costs we have not explicitly taken into account the cost to return the response back to the querying node. For the structured scheme, this is easy to incorporate as the response is returned along the reverse path as the directed query, and hence incorporating this cost is equivalent to simply doubling the cost (which can be absorbed into the constant term). For the unstructured scheme, the cost of a directed response will be of the order $O(\frac{\sqrt{N}}{\sqrt{r}})$ and hence, for the large networks that are the focus of this study, negligible compared to the $O\left(\frac{N}{r+1}\right)$ cost of the blind search.

Looking at equations (1), (2), and (3), we find that, as expected, the replication costs increase with the number of replicas, while the search costs decrease with the number of replicas. We can resolve this tradeoff by considering the aggregate total expected cost of search and querying and optimizing for it.

	c_1	c_2	c_3
Grid	$\frac{2}{3}$	1	1
Rand.	$\frac{0.52}{R\sqrt{\rho}}$	$\frac{c}{R\sqrt{\rho}}$	2.15

Table 1: The constants for the cost expressions (1), (2), and (3) for both regular grid and uniform random deployments, where R is the radio radius defined in section 2, c is a constant in $(0.66, 1.71)$, and ρ is the density of nodes.

The following is the common form of the total cost:

$$C_t = \sum_{i=1}^{m} q_i C_s(r_i) + \sum_{i=1}^{m} C_r(r_i) \tag{4}$$

where $C_s(r_i)$ is the expected search cost of i^{th} event and $C_r(r_i)$ is its expected replication cost.

From the above, we get the following expressions for the expected total energy cost for all events which consists of search costs weighed by the number of queries as well as the replication costs:

1. Under the unstructured replication scheme, the total energy cost is

$$C_{tot,u} = \sum_{i=1}^{m} c_2 \frac{N q_i}{r_i+1} + \sum_{i=1}^{m} c_1 \sqrt{N}(r_i-1) \tag{5}$$

2. Under the structured replication scheme

$$C_{tot,s} = \sum_{i=1}^{m} c_3 \frac{\sqrt{N} q_i}{\sqrt{r_i}} + \sum_{i=1}^{m} c_1 \sqrt{N}(r_i-1) \tag{6}$$

To simplify our expressions, with a slight abuse of notation, we shall make the following substitutions: in equation (7), after dividing both sides by c_1, we let $C_{tot,u}/c_1 \to C_{tot,u}$ and $\frac{c_2}{c_1} q_i \to q_i$; in equation (8), after dividing both sides by c_1, we let $C_{tot,s}/c_1 \to C_{tot,s}$ and $\frac{c_3}{c_1} q_i \to q_i$. And the following expressions are the simplified versions;

$$C_{tot,u} = \sum_{i=1}^{m} \frac{N q_i}{r_i+1} + \sum_{i=1}^{m} \sqrt{N}(r_i-1) \tag{7}$$

$$C_{tot,s} = \sum_{i=1}^{m} \frac{\sqrt{N} q_i}{\sqrt{r_i}} + \sum_{i=1}^{m} \sqrt{N}(r_i-1) \tag{8}$$

4. OPTIMIZATION FORMULATION

Now we can formulate the problem of optimizing the total cost as follows;

$$\begin{aligned} &\textit{Minimize} \quad C_t = \textstyle\sum_{i=1}^{m} q_i C_s(r_i) + \sum_{i=1}^{m} C_r(r_i) \\ &\textit{s.t} \quad \textstyle\sum_{i=1}^{m} r_i \leq S \end{aligned} \tag{9}$$

The optimization formulation does require global knowledge of query rates for each event and hence the optimum may not be necessarily achieved by distributed heuristics in practice, but this is still a useful tool for our investigations of performance scalability as it provides the best-case scenario. We solve this problem using the method of Lagrange multipliers. The Lagrangian function for this inequality-constrained optimization problem can be expressed using a Lagrange multiplier λ and a slack variable x as follows;

$$L(\bar{r}, \lambda, x) = C_t + \lambda(\sum_{i=1}^{m} r_i - S + x^2) \tag{10}$$

It can be shown that the objective functions for both the unstructured and structured scheme are all convex. Thus, first-order conditions are sufficient for global optimization. Solving these conditions, we find that

i) When the constraint is inactive (i.e. $\lambda = 0$), we have that

$$r^*_{i,inact} = \begin{cases} q_i^{1/2} N^{1/4} - 1, & \text{(unstructured)} \\ \beta_s \cdot q_i^{2/3}, & \text{(structured)} \end{cases} \tag{11}$$

where

$$\beta_s = 2^{-2/3} \tag{12}$$

ii) When the constraint is active (i.e. $x = 0$, $\lambda \geq 0$), we get

$$r^*_{i,act} = \begin{cases} \frac{S+m}{\sum_{j=1}^m \sqrt{q_j}} \sqrt{q_i} - 1, & \text{(unstructured)} \\ \frac{S}{\sum_{j=1}^m q_j^{2/3}} \, q_i^{2/3}, & \text{(structured)} \end{cases} \tag{13}$$

Now we can derive the optimal expected total energy costs substituting equation (11) and (13) into equation (7) and (8) respectively as follows;

i) For the unstructured network

$$C^*_{t,u} = \begin{cases} \sum_{i=1}^m \sqrt{N}\left(N^{1/4}\sqrt{q_i} - 2\right) \\ \quad + \sum_{i=1}^m N^{3/4}\sqrt{q_i}\,, & \text{(Inactive)} \\ \sum_{i=1}^m \sqrt{N}\left(\frac{(m+S)}{\sum_{j=1}^m \sqrt{q_j}}\sqrt{q_i} - 2\right) \\ \quad + \sum_{i=1}^m \frac{\sum_{j=1}^m \sqrt{q_j}}{m+S}\sqrt{q_i}\, N, & \text{(Active)} \end{cases} \tag{14}$$

ii) For the structured network

$$C^*_{t,s} = \begin{cases} \sum_{i=1}^m \frac{1}{\sqrt{\beta_s}}\sqrt{N} q_i^{2/3} \\ \quad + \sum_{i=1}^m \sqrt{N}\left(\beta_s\, q_i^{2/3} - 1\right), & \text{(Inactive)} \\ \sum_{i=1}^m \sqrt{N}\left(\frac{S}{\sum_{j=1}^m q_j^{2/3}}\, q_i^{2/3} - 1\right) \\ \quad + \sum_{i=1}^m \frac{\sqrt{\sum_{j=1}^m q_j^{2/3}}}{\sqrt{S}}\, q_i^{2/3}\sqrt{N}, & \text{(Active)} \end{cases} \tag{15}$$

In order to have better understanding in the behavior of the optimal total cost, we look into optimal total costs assuming that the query rate for each item is same one another, that is $q_i = q, \forall\, i$. Figure 1 shows the optimal per-node total cost (which equals the optimal total cost divided by the number of nodes N) vs. the number of events (m) as X axis and the query rate (q) as Y axis when $N = 10^4$. The curved thick line represents the boundary of enough storage for unconstraint optimal point. Beyond that boundary, the surface increases sharply and it is more sensitive to the increase in the number of events than that of query rate. Note that the structured replication scheme has a gentler incline and larger unconstrained region than the unstructured replication scheme.

5. SCALING CONDITIONS FOR BOUNDED STORAGE

As we have seen above, when the available storage in the network exceeds the sum of the unconstrained optimum number of copies for all events, we have an efficient region where the network can achieve the smallest total energy cost of querying (and replication). From a scalability perspective, it is desirable to ensure that the per-node storage requirements remain bounded irrespective of the network size. This is equivalent to requiring that there be a constant storage s per node such that the total storage $S = s \cdot N$.

DEFINITION 1. *We say that a network* ***scales efficiently with bounded storage*** *if*

$$\exists N_0 \;\; s.t. \;\; \forall N > N_0, \;\; \sum_{i=1}^m r^*_{i,inact} < S = s \cdot N \tag{16}$$

To obtain useful insights regarding scalability, we simplify our expressions from this point on by assuming that the query rate for all events is uniform, i.e., $q_i = q, \forall i$. We now give scaling results that quantify the above condition for structured and unstructured networks.

THEOREM 1. ***Conditions for Efficient Operation of Unstructured Networks with Bounded Storage:*** *For unstructured networks, if condition (16) holds, then $m \cdot q^{1/2}$ must be $O\left(N^{3/4}\right)$. Further, if $m \cdot q^{1/2}$ is $o\left(N^{3/4}\right)$, then condition (16) holds.*

Proof: If condition (16) holds, then the following holds for all $N > N_0$:

$$\sum_{i=1}^m r^*_{i,inact} \; = \; m\, q^{1/2} N^{1/4} - m \; \leq \; sN$$

$$\Rightarrow \quad m\,(q^{1/2} - N^{-1/4}) \; \leq \; s\, N^{3/4} \tag{17}$$

$$\Rightarrow \quad m\, q^{1/2} \; \leq \; s\, N^{3/4} \tag{18}$$

Since s is constant, $m\, q^{1/2}$ is $O\left(N^{3/4}\right)$. Note that inequality (18) holds for the sufficiently large $N > N_0$ since $N^{-1/4}$ goes to zero, as N goes to infinity.

On the other hand, if $m\, q^{1/2}$ is $o\left(N^{3/4}\right)$, then for $N > N_0$ and any arbitrary small positive constant ϵ,

$$m\, q^{1/2} < \epsilon N^{3/4} \leq s N^{3/4}$$

$$\Rightarrow \quad m\, q^{1/2} - m\, N^{-1/4} < m\, q^{1/2} < sN^{3/4}$$

$$\Rightarrow \quad m\left(q^{1/2}N^{1/4} - m\right) = \sum_{i=1}^m r^*_{i,inact} < sN = S$$

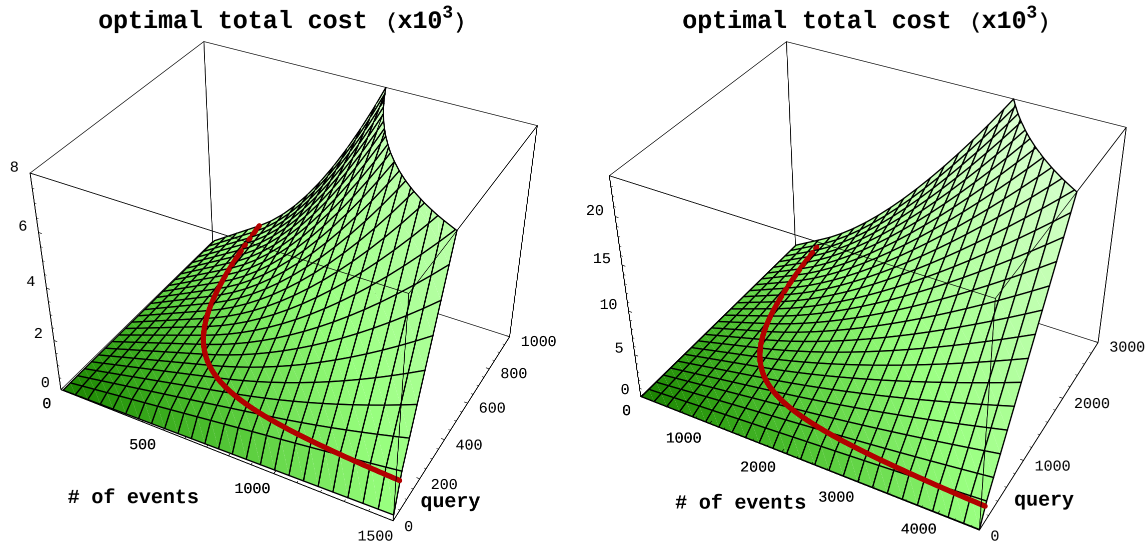
□

THEOREM 2. ***Conditions for Efficient Operation of Structured Networks with Bounded Storage:*** *For structured networks, if condition (16) holds, then $q^{2/3} \cdot m$ must be $O(N)$. Further, if $q^{2/3} \cdot m$ is $o(N)$, then condition (16) holds.*

Proof: It can be proved in the same way as proof of Theorem 1 using the structured case of equation (11).

□

Theorem 1 and Theorem 2 are not symmetric. It is important to note that it is possible that the network is operating inefficiently in the constrained region when the $q^{1/2} \cdot m$ is $\Theta\left(N^{3/4}\right)$ (in case of



(a) (b)

Figure 1: (a) The per-node total cost of the unstructured network of $N = 10000$ (b) that of the structured

unstructured networks), and $q^{2/3} \cdot m$ is $\Theta(N)$ (in case of structured networks).

To understand the implications of these theorems, it is helpful to consider some extreme cases of the scaling behavior of the number of events (m) and the query rate (q). We consider allowing each of these parameters to scale as $\Theta(1)$ or $\Theta(N)$, giving us four possible combinations. In practice the scaling behavior of the events and queries with network size is determined by the application scenario. For instance, an application which requires the network (regardless of its size) to have only a single sink injecting queries for events would have that q is $\Theta(1)$, while a richer application involving increasing numbers of users with the network size could have that $\Theta(N)$. For many event monitoring applications, it is likely to be reasonable to assume that the number of observed events scales proportionally with the deployment area which for a constant density deployment would mean that m is $\Theta(N)$; however in other applications the scaling of m may be weaker, all the way down to the extreme of $\Theta(1)$ (which would imply that there only a finite number of events that can be detected regardless of the network size).

Consider each combination first for the case of unstructured networks. When q and m are both $\Theta(1)$, then by Theorem 1, in this case the networks always scale with bounded storage; when q and m are both $\Theta(N)$, then $q^{1/2} \cdot m$ becomes $\Theta\left(N^{3/2}\right)$ and hence (since this is not $O\left(N^{3/4}\right)$), by Theorem 1, the network never scales with bounded storage. The following table summarizes the scalability for each case with unstructured networks:

q \ m	$\Theta(1)$	$\Theta(N)$
$\Theta(1)$	Always	Never
$\Theta(N)$	Always	Never

Table 2: Illustration of scenarios under which unstructured networks scale efficiently with bounded storage.

Similarly, we can apply Theorem 2 to analyze the scalability for structured networks for these illustrative scenarios. This is summarized in the following table. Here, one interesting case is that when q is $\Theta(1)$ and m is $\Theta(N)$, the network can potentially operate in either the active storage constraint or the inactive storage constraint region as it scales. This is because in this case $q^{2/3} \cdot m$ is $\Theta(N)$, so that the second (efficiency-guaranteeing) clause of Theorem 2 does not apply.

q \ m	$\Theta(1)$	$\Theta(N)$
$\Theta(1)$	Always	Possibly
$\Theta(N)$	Always	Never

Table 3: Illustration of scenarios under which structured networks scale efficiently with bounded storage.

6. THE SCALING BEHAVIOR OF ENERGY COSTS

We now examine the scaling behavior of the total energy costs for both unstructured and structured networks.

THEOREM 3. *The total energy costs for unstructured networks grow with network size N as follows:*

$$C^*_{t,u} = \begin{cases} \Theta\left(m \cdot q^{1/2} \cdot N^{3/4}\right), & \textit{(inactive)} \\ \Theta\left(N^{3/2} + m^2 \cdot q\right), & \textit{(active)} \end{cases} \tag{19}$$

Proof: In the inactive constraint region, the total energy cost is given from equation (14) by,

$$\begin{aligned} &\sum_{i=1}^{m} \sqrt{N}\left(N^{1/4}\sqrt{q_i} - 2\right) + \sum_{i=1}^{m} N^{3/4}\sqrt{q_i} \\ &= \; 2mq^{1/2}N^{3/4} - 2mN^{1/2} \\ &= \; \Theta\left(mq^{1/2}N^{3/4}\right) \end{aligned}$$

In the active constraint region, the total energy cost is given from equation (14) by,

$$\begin{aligned}
&\sum_{i=1}^{m}\sqrt{N}\left(\frac{(m+S)}{\sum_{j=1}^{m}\sqrt{q_j}}\sqrt{q_i}-2\right)\\
&\quad+\sum_{i=1}^{m}\frac{\sum_{j=1}^{m}\sqrt{q_j}}{m+S}\sqrt{q_i}\,N\\
&= sN^{3/2}-N^{1/2}m+\frac{m^2qN}{m+sN}\\
&= \Theta\left(N^{3/2}+m^2q\right) \qquad (20)
\end{aligned}$$

Since it is reasonable to consider that the number of events m is at most proportional to N, sN is dominant compared to m. Thus, $\frac{m^2qN}{m+sN}$ is $\Theta\left(m^2q\right)$, and so equation (20) holds.

□

THEOREM 4. *The total energy costs for structured networks grow with network size N as follows:*

$$C^*_{t,s}=\begin{cases}\Theta\left(m\cdot q^{2/3}\cdot N^{1/2}\right), & \textit{(inactive)}\\ \Theta\left(N^{3/2}+m^{3/2}\cdot q\right), & \textit{(active)}\end{cases} \qquad (21)$$

Proof: It can be proved in the same way as the proof of Theorem 3 using the equation (15)

□

To illustrate the scaling of these costs, we again consider the four scenarios pertaining to q and m. As we observed in Table 2, for the unstructured networks, scaling with unbounded storage is observed only when m is $\Theta(1)$ (regardless of q); when m is $\Theta(N)$, then the network operates in the active constraint region as it scales. Substituting into the relevant cases of Theorem 3, therefore, we get the following table for the four cases.

q \ m	$\Theta(1)$	$\Theta(N)$
$\Theta(1)$	$\Theta(N^{3/4})$	$\Theta(N^2)$
$\Theta(N)$	$\Theta(N^{5/4})$	$\Theta(N^3)$

Table 4: Illustration of the scaling of total energy costs for unstructured networks.

We generate a similar table below using Theorem 4 to illustrate the scenarios for structured networks. As mentioned above, when m is $\Theta(N)$ and q is $\Theta(1)$, both active and inactive constraint regions are possible. However, it turns out that in both cases the scaling shows the same order ($\Theta\left(N^{3/2}\right)$).

q \ m	$\Theta(1)$	$\Theta(N)$
$\Theta(1)$	$\Theta(N^{1/2})$	$\Theta(N^{3/2})$
$\Theta(N)$	$\Theta(N^{7/6})$	$\Theta(N^{5/2})$

Table 5: Illustration of the scaling of total energy costs for structured networks.

We observe something striking about Tables 4 and 5. In both tables, among the four cases, only when both q and m are $\Theta(1)$ do we observe that the total costs for the whole network scale as $O(N)$. In other words, only in this example case do we have $O(1)$ scaling of the per-node cost, i.e. bounded energy consumption per node. This motivates us to inquire about the general conditions under which a network can scale while ensuring that the energy requirement per node is kept bounded — a very important requirement from a practical perspective.

THEOREM 5. *For unstructured networks, the energy requirement per node is bounded if and only if*

$$q^{1/2}\cdot m \;\text{ is }\; O\left(N^{1/4}\right)$$

Proof: the total optimal energy cost per node is the total cost divided by the number of nodes N. If the energy requirement per node is bounded, the per-node total energy cost must be $O(1)$. From Theorem 3, the per-node total cost cannot be bounded regardless of m and/or q in the active constraint region since it is at least $\Theta(N^{1/2})$. In the inactive constraint region, however, the per-node total cost is given from equation (14) divided by N (assuming $q_i = q,\ \forall i$) as follows:

$$C^*_{t,u}/N \;=\; 2mq^{1/2}N^{-1/4}-2mN^{-1/2} \;\le\; \mathcal{C}_0$$

where $\mathcal{C}_0$ is a sufficiently large constant.

$$\begin{aligned}
&\Rightarrow\quad m\cdot\left(q^{1/2}-N^{-1/4}\right)\le\frac{\mathcal{C}_0}{2}N^{1/4}\\
&\Rightarrow\quad mq^{1/2}\le\frac{\mathcal{C}_0}{2}N^{1/4},\quad \forall N>N_0 \qquad (22)
\end{aligned}$$

Note that inequality (22) holds since $N^{-1/4}$ goes zero as N goes to infinity. Therefore, $mq^{1/2}$ is $O\left(N^{1/4}\right)$.

On the other hand, if $mq^{1/2}$ is $O\left(N^{1/4}\right)$,

$$\begin{aligned}
&mq^{1/2}\;\le\;\mathcal{C}_0N^{1/4}\\
\Rightarrow\;& m\cdot\left(q^{1/2}-N^{-1/4}\right)\;\le\;\mathcal{C}_0\,N^{1/4}\\
\Rightarrow\;& 2mq^{1/2}N^{-1/4}-2mN^{-1/2}\le\mathcal{C}_0/2 \qquad (23)
\end{aligned}$$

Note that the left side of inequality (23) is equal to the optimized per-node total energy cost in the inactive constraint region. As for the total cost in the active constraint region, however, since the assumption that is $mq^{1/2}$ is $O\left(N^{1/4}\right)$ already satisfies the condition of theorem 1, it is sufficient to consider the total cost in the inactive constraint region only. Therefore, the per-node total energy cost is bounded as N goes to infinity.

□

THEOREM 6. *For structured networks, the energy requirement per node is bounded if and only if*

$$q^{2/3}\cdot m \;\text{ is }\; O\left(N^{1/2}\right)$$

Proof: It can be proved in the same way as the proof of theorem 5 using the equation (15).

□

COROLLARY 1. *For both structured and unstructured networks, if the energy requirement per node is bounded, the networks also scale with bounded storage. i.e., the bounded energy requirement is a stricter condition than scaling with bounded storage.*

7. NETWORK SCALING ON FIXED ENERGY BUDGET

So far, we have seen the conditions for bounded storage and energy and the scaling of energy costs as a function of event and activity rates. We now consider having a fixed energy budget, and look into what conditions the network size must satisfy to ensure that events and queries within the finite deployment time duration can be resolved before energy depletion. Specifically, we will assume that there is an average energy budget e for each node, so that the total energy is $E = e\cdot N$.

DEFINITION 2. *We say a network **operates successfully** if it can satisfy all queries for all events in a given deployment period before energy depletion. This requires that $C_t \leq e \cdot N$.*

THEOREM 7. *For unstructured networks, given fixed average per-node energy e (i.e., the total energy allocated optimally among the nodes in the network grows linearly with the network size as $E = e \cdot N$), the following statements describe the conditions on the network size N, the number of events m and the number of queries per event q that ensure that the network can be operated successfully.*

1. *If $m \cdot q^{1/2}$ is $o(N^{1/4})$, then there exists a minimum network size $N_{min}(e)$ beyond which it can always be operated successfully.*

2. *If $m \cdot q^{1/2}$ is $\Theta(N^{1/4})$, then there exists an average per-node energy e^* such that for all $e < e^*$, it is not possible to operate a network of any size successfully, while for all $e \geq e^*$ it is possible to operate a network of any size successfully.*

3. *If $m \cdot q^{1/2}$ is $\omega(N^{1/4})$, but $o(N)$, then there exists a maximum network size $N_{max}(e)$ beyond which the network cannot be operated successfully. Further N_{max} is a convex function of e.*

4. *If $m \cdot q^{1/2}$ is $\Theta(N)$, then there exists a maximum network size $N_{max}(e)$ beyond which the network cannot be operated successfully. Further N_{max} increases linearly with e.*

5. *If $m \cdot q^{1/2}$ is $\omega(N)$, then there exists a maximum network size $N_{max}(e)$ beyond which the network cannot be operated successfully. Further N_{max} increases as a concave function of e.*

Proof:

1. $m \cdot q^{1/2} = \Theta(N^{1/4-\epsilon})$ where $\epsilon > 0$. Then, the optimal total cost is given from Theorem 3 by,

$$\begin{aligned} C^*_{t,u,inactive} &= \Theta(m \cdot q^{1/2} N^{3/4}) = \Theta(N^{1-\epsilon}) \\ &= \alpha N^{1-\epsilon} + o(N^{1-\epsilon}) \end{aligned}$$

Since the total cost expenditure should be less than the given energy $e \cdot N$,

$$\alpha N^{1-\epsilon} + o(N^{1-\epsilon}) \leq eN$$

Note that there exists $N \geq N_0$ such that this inequality holds, where N_0 is a fixed constant and can be considered as the minimum network size to make the network operate successfully. Note that this condition satisfies the theorem 1 and so the network is in the inactive constraint region.

2. We can prove this case in the same way as the case 1.f $m \cdot q^{1/2} = \Theta(N^{1/4})$. Then, the total cost is given by,

$$\begin{aligned} C^*_{t,u,inactive} &= \Theta(m \cdot q^{1/2} N^{3/4}) = \Theta(N) \\ &= \alpha N + o(N) \end{aligned}$$

From the total cost expenditure constraints,

$$\alpha N + o(N) \leq eN$$

Note that there exists $e \geq e^* > \alpha$ such that this inequality holds for all N.

3. In this case, we have two sub-cases. If $m \cdot q^{1/2} = O(N^{3/4})$, we should use the corresponding inactive cost by Theorem 1. Otherwise, we should use the active cost. First of all, let's consider the first sub-case. $m \cdot q^{1/2} = \Theta(N^{1/4+\epsilon})$ where $0 < \epsilon \leq 1/2$. Then the optimum total cost is given by,

$$\begin{aligned} C^*_{t,u,inactive} &= \Theta(m \cdot q^{1/2} N^{3/4}) = \Theta(N^{1+\epsilon}) \\ &= \alpha N^{1+\epsilon} + o(N^{1+\epsilon}) \end{aligned}$$

From the the total cost expenditure constraints,

$$\alpha N^{1+\epsilon} + o(N^{1+\epsilon}) \leq eN$$

Note that there exists N_{max} such that it achieves the equality. For $e >> \alpha$, N_{max} can be approximated as follows:

$$N_{max} = (1/\alpha)^{1/\epsilon} \cdot e^{1/\epsilon}$$

, where $1/\epsilon \geq 2$. Therefore, this N_{max} is a convex function of e.

Now, let's consider the second sub-case, where $m \cdot q^{1/2} = \Theta(N^{3/4+\epsilon})$, $0 \leq \epsilon < 1/4$ and we should use the active total cost. Through the similar reasoning, we can easily achieve the following equality with approximation for $e >> \alpha$.

$$N_{max} = (\frac{1}{\alpha})^{\frac{2}{4\epsilon+1}} \cdot e^{\frac{2}{4\epsilon+1}}$$

where $1 < \frac{2}{4\epsilon+1} \leq 2$. Therefore, this N_{max} is a convex function of e.

As for cases 4 and 5, they can be proved in the same way as case 3 using the active total cost equation.

□

THEOREM 8. *For structured networks, given fixed average per-node energy e (i.e., the total energy allocated optimally among the nodes in the network grows linearly with the network size as $E = e \cdot N$), the following statements describe the conditions on the network size N, the number of events m and the number of queries per event q that ensure that the network can be operated successfully.*

1. *If $m \cdot q^{2/3}$ is $o(N^{1/2})$, then there exists a minimum network size $N_{min}(e)$ beyond which it can always be operated successfully.*

2. *If $m \cdot q^{2/3}$ is $\Theta(N^{1/2})$, then there exists an average per-node energy e^* such that for all $e < e^*$, it is not possible to operate a network of any size successfully, while for all $e \geq e^*$ it is possible to operate a network of any size successfully.*

3. *If $m \cdot q^{2/3}$ is $\omega(N^{1/2})$, but $o(N^{4/3})$, then there exists a maximum network size $N_{max}(e)$ beyond which the network cannot be operated successfully. Further N_{max} is a convex function of e.*

4. *If $m \cdot q^{2/3}$ is $\Theta(N^{4/3})$, then there exists a maximum network size $N_{max}(e)$ beyond which the network cannot be operated successfully. Further N_{max} increases linearly with e.*

5. *If $m \cdot q^{2/3}$ is $\omega(N^{4/3})$, then there exists a maximum network size $N_{max}(e)$ beyond which the network cannot be operated successfully. Further N_{max} increases as a concave function of e.*

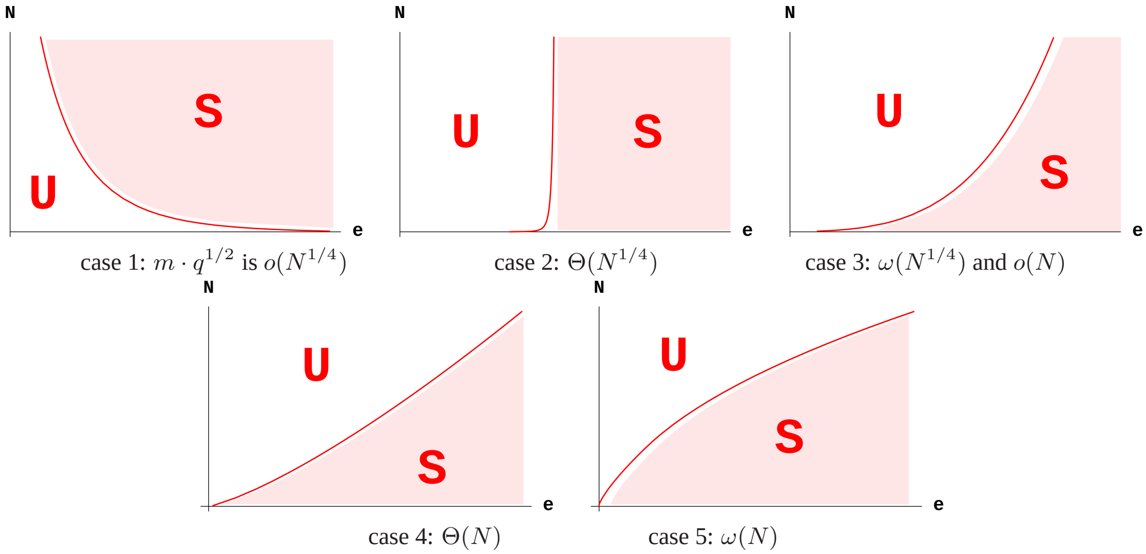


Figure 2: Network size conditions for successful operation with respect to per-node energy budget for different event-rate and query-rate scaling behaviors, for an unstructured network; S denotes the successful region while U denotes the unsuccessful region.

Proof: This is similar to the proof for Theorem 7 □

Figure 2 illustrates the network size versus energy budget curves for the five different cases in Theorem 7. It is obtained numerically by equating the expressions for total cost with the energy budget $E = e \cdot N$, and solving for N as a function of e, under particular m and q scaling settings that satisfy each of the corresponding cases. (A very similar figure can be obtained for structured networks and is omitted due to lack of space). The regions marked S and U are where the network operates successfully and unsuccessfully, respectively.

We see that under case 1, there is a minimum network size that is needed to ensure successful operation, and this minimum network size decreases rapidly with increasing energy availability. In this case, the event and query activity remains low enough that adding nodes to the network is beneficial (as it increases the available total energy). Under the event-query activity case 2, there exists an per-node energy threshold such that below this threshold, no network can operate successfully, but beyond this threshold, networks of any size can be operated. Under cases 3, 4, and 5, we see that for a given energy budget there exist maximum network sizes beyond which successful operation is impossible. In these cases, adding nodes to the network is harmful as each additional node introduces more consumption than resources. The key distinction between these cases is that under case 3, there is a convex growth that implies that adding energy resources to each node provides a super-linear improvement in the maximum network size that can be sustained; under case 4, the maximum network size grows linearly with the per-node energy budget; and under case 5, the concave growth of the curve implies that adding energy resources provide diminishing returns in maximum network size.

8. NETWORK LIFETIME SCALING

We now consider a relaxation of one of our key assumptions — that the network is being operated for a fixed duration. This allows us to examine how the lifetime of the network (the period over which all queries for all events can be resolved successfully) scales with the network size. In this connection we will assume that the total number of events since network initiation and the total number of queries per event $(m(t), q(t))$ are such that they are both non-decreasing functions of time, and at least one is a strictly increasing function of time.

THEOREM 9. *For unstructured networks, with a fixed average per-node energy budget of e, so long as the number of events and queries scale temporally so that $m \cdot q^{1/2}$ is an increasing function of time, the lifetime of deployment T over which the network can operate successfully scales with the network size as per the following conditions:*

1. *if $m \cdot q^{1/2}$ is $o(N^{1/4})$ then T increases with N.*
2. *if $m \cdot q^{1/2}$ is $\Theta(N^{1/4})$ then T is constant with respect to N.*
3. *if $m \cdot q^{1/2}$ is $\omega(N^{1/4})$ then T decreases with N.*

Proof:

1. $m \cdot q^{1/2} = \Theta(N^{1/4-\epsilon} \cdot T^{\beta})$, where $\epsilon > 0, \beta > 0$. Then, the optimal total cost is given from Theorem 1 by,

$$C^*_{t,u,inact} = \Theta(m \cdot q^{1/2} \cdot N^{3/4}) = \Theta(N^{1-\epsilon} \cdot T^{\beta}) = \alpha N^{1-\epsilon} T^{\beta} + o(N^{1-\epsilon} T^{\beta})$$

   From the total cost expenditure constraints,

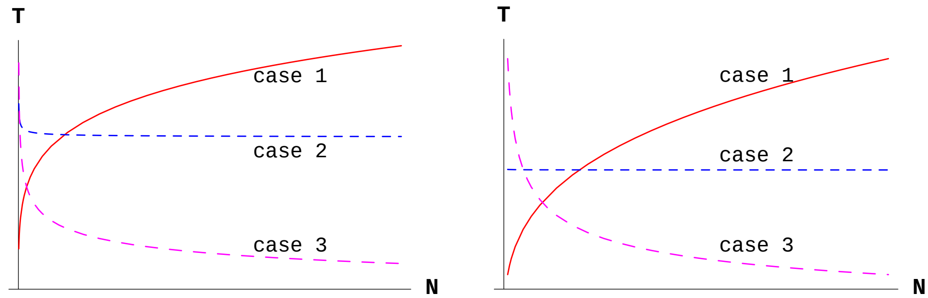
$$\alpha N^{1-\epsilon} T^{\beta} + o(N^{1-\epsilon} T^{\beta}) \leq eN$$

   Note that there exists T_{max} such that it satisfies the above equality; $T < T_{max}$ satisfies the inequality. For $e >> \alpha$, T_{max} can be approximated as follows:

$$T_{max} = \left(\frac{e}{\alpha}\right)^{1/\beta} \cdot N^{\epsilon/\beta}$$

   where $\frac{\epsilon}{\beta} > 0$.
   Therefore, this T_{max} increases with N.

2. We can prove this in the similar way as the case 1. $m \cdot q^{1/2} = \Theta(N^{1/4} T^{\beta})$, where $\beta > 0$. Then, the optimal total cost is



(a) (b)

Figure 3: The network lifetime (T) vs. the number of nodes (N) of the (a) unstructured network and (b) the structured networks when both m and q are proportional to T

given by,

$$\begin{aligned} C^*_{t,u,inact} &= \Theta(m \cdot q^{1/2} \cdot N^{3/4}) = \Theta(N \cdot T^\beta) \\ &= \alpha N T^\beta + o(NT^\beta) \end{aligned}$$

From the total cost expenditure constraints,

$$\alpha N T^\beta + o(NT^\beta) \leq eN$$

Note that there exists T_{max} such that it satisfies the above equality. Further, for $e >> \alpha$, T_{max} can be approximated as follows:

$$T_{max} = \left(\frac{e}{\alpha}\right)^{1/\beta}$$

Therefore, this T_{max} is constant with respect to N.

3. As the case 3 of Theorem 7, we also have two sub-cases here. First of all, consider $m \cdot q^{1/2} = \Theta(N^{1/4+\epsilon}T^\beta)$, where $0 < \epsilon \leq 1/2, \beta > 0$. Then, the optimal total cost is given by,

$$\begin{aligned} C^*_{t,u,inact} &= \Theta(m \cdot q^{1/2} \cdot N^{3/4}) \\ &= \Theta(N^{1+\epsilon} \cdot T^\beta) \\ &= \alpha N^{1+\epsilon} T^\beta + o(N^{1+\epsilon}T^\beta) \end{aligned}$$

From the total cost expenditure constraints,

$$\alpha N^{1+\epsilon} T^\beta + o(N^{1+\epsilon}T^\beta) \leq eN$$

Note that there exists T_{max} such that it satisfies the above equality. Further, for $e >> \alpha$, T_{max} can be approximated as follows:

$$T_{max} = \left(\frac{e}{\alpha}\right)^{1/\beta} N^{-\epsilon/\beta}$$

where $-\epsilon/\beta < 0$. Therefore, this T_{max} decreases with N.

Now, let's consider the second sub-case, where $m \cdot q^{1/2} = \Theta(N^{3/4+\epsilon}T^\beta)$ with $\epsilon > 1/2, \beta > 0$ and we should use the active total cost. Through the similar reasoning, we can easily achieve the following equality for $e >> \alpha$:

$$T_{max} = \left(\frac{e}{\alpha}\right)^{\frac{1}{2\beta}} N^{-\frac{1}{4\beta}-\frac{\epsilon}{\beta}}$$

where $-\frac{1}{4\beta} - \frac{\epsilon}{\beta} < 0$. Therefore, this T_{max} decreases with N.

□

THEOREM 10. *For structured networks, with a fixed average per-node energy budget of e, so long as the number of events and queries scale temporally so that $m \cdot q^{2/3}$ is an increasing function of time, the lifetime of deployment T over which the network can operate successfully scales with the network size as per the following conditions:*

1. *if $m \cdot q^{2/3}$ is $o(N^{1/2})$ then T increases with N.*
2. *if $m \cdot q^{2/3}$ is $\Theta(N^{1/2})$ then T is constant with respect to N.*
3. *if $m \cdot q^{2/3}$ is $\omega(N^{1/2})$ then T decreases with N.*

Proof: This is similar to the proof for Theorem 9.

□

These theorems are illustrated in Figure 3 through a numerical plot based on exact expressions. We can see that event-query scaling conditions determine whether the lifetime of the deployed network increases, decreases, or remains constant with respect to network size.

9. CONCLUSIONS AND FUTURE WORK

We have investigated the fundamental scaling behavior of storage and querying in wireless sensor networks. The main take away from this study is that the event and query rates must scale sufficiently slowly with the network size if scalable performance is desired. In particular, an important scaling condition is ensuring that $q^{1/2} \cdot m$ be $O(N^{1/4})$ for unstructured networks, and that $q^{2/3} \cdot m$ be $O(N^{1/2})$ for structured networks. Satisfying this condition ensures that adding nodes to the network is beneficial in that the energy and storage resources they bring outweigh the additional event and query activity they induce. This can be seen from many perspectives: satisfying this condition implies that (i) sensor networks require bounded energy and storage per node, (ii) arbitrarily large networks can be operated successfully with a limited energy budget, and (iii) that the network lifetime increases with network size for a given energy budget.

In our study we have not explicitly considered bandwidth capacity; we have implicitly assumed that the energy constraints will be more severe than bandwidth constraints in the system. However, if energy constraints are not significant (consider as an extreme case if all nodes could be wired for power), bandwidth issues could be the dominant consideration. This is a topic for future work.

We have made the strong assumption that queries are uniformly distributed. However, our results showing the existence of a maximum network size for a given energy budget can be potentially interpreted as an argument that queries need to be kept localized to within a fixed distance of corresponding events. In a practical large-scale system where queries are uniformly generated and the rate of events and queries large enough that the scalability thresholds are exceeded, these results motivate the decomposition of large-scalable sensor networks into a two-tier architecture. In this case, the lower-tier would consist of the wireless nodes within each limited-size cluster, while the upper-tier would provide a wired connection between cluster-heads that can be used to inject queries from any point in the network into any cluster with minimal energy expense.

In the future, we would like to explicitly consider scalability under localized queries. We would also like to undertake realistic simulations and large-scale experiments to validate the analytical results presented in this work.

10. REFERENCES

[1] R. Govindan, “Data-centric Routing and Storage in Sensor Networks,” in *Wireless Sensor Networks*, Eds. Raghavendra, C.S., Sivalingam, K.M., and Znati, T., Kluwer Academic Publishers, 2004, pp. 185-206.

[2] S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, R. Govindan, L. Yin, and F. Yu, “Data-Centric Storage in Sensornets with GHT, A Geographic Hash Table”, In *Mobile Networks and Applications (MONET), Special Issue on Wireless Sensor Networks*, 8:4, Kluwer, August 2003.

[3] N. Chang and M. Liu, “Revisiting the TTL-based Controlled Flooding Search: Optimality and Randomization,” *Proceedings of the Tenth Annual International Conference on Mobile Computing and Networks (ACM MobiCom)*, September, 2004.

[4] M. Chu, H. Haussecker, F. Zhao, “Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks,” *International Journal of High Performance Computing Applications*, vol. 16, no. 3, pp. 293 - 313, August 2002.

[5] B. Krishnamachari, J. Ahn, “Optimizing Data Replication for Expanding Ring-based Queries in Wireless Sensor Networks”, *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WIOPT ’06*), April 2006, Boston, USA.

[6] J. Ahn and B. Krishnamachari, “Derivations of the Expected Energy Costs of Search and Replication in Wireless Sensor Networks”, *USC Computer Engineering Technical Report CENG-2006-3*, April 2006.

[7] P. Gupta and P. R. Kumar, “The Capacity of Wireless Networks,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, March 2000.

[8] Z. Hu and B. Li, “Fundamental Performance Limits of Wireless Sensor Networks,” *Ad Hoc and Sensor Networks*, Yang Xiao and Yi Pan, Editors, Nova Science Publishers, 2004.

[9] M. Bhardwaj and A. P. Chandrakasan, “Bounding the Lifetime of Sensor Networks Via Optimal Role Assignments,” *Proceedings of IEEE INFOCOM*, June 2002.

[10] J.-H. Chang, L. Tassiulas, “Maximum Lifetime Routing In Wireless Sensor Networks,” *IEEE/ACM Transactions on Networking*, vol. 12, no. 4, pp. 609-619, August 2004.

[11] K. Kalpakis, K. Dasgupta, and P. Namjoshi, “Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks,” *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 42, no. 6, pp. 697-716, August 2003.

[12] S. Shenker, S. Ratnasamy, B. Karp, R. Govindan, and D. Estrin, “Data-Centric Storage in Sensornets”, ACM SIGCOMM, Computer Communications Review, Vol. 33, Num. 1, January 2003.

[13] N. Trigoni, Y. Yao, A. Demers, J. Gehrke, and R. Rajaraman. “Hybrid Push-Pull Query Processing for Sensor Networks”, In Proceedings of the Workshop on Sensor Networks as part of the GI-Conference Informatik 2004. Berlin, Germany, September 2004.

[14] S. Shakkottai, Asymptotics of Query Strategies over a Sensor Network, INFOCOM’04, March 2004

[15] X. Liu, Q. Huang, Y. Zhang, “Combs, Needles, Haystacks: Balancing Push and Pull for Discovery in Large-Scale Sensor Networks”, ACM Sensys, November 2004

[16] S. Kapadia and B. Krishnamachari, “Comparative Analysis of Push-Pull Query Strategies for Wireless Sensor Networks,” International Conference on Distributed Computing in Sensor Systems (DCOSS), June 2006.