Data Acquisition in Multiple-sink Sensor Networks

Abhimanyu Das, Debojyoti Dutta
University of Southern California, Los Angeles, CA 90089, USA
abhimand@usc.edu,ddutta@usc.edu

I. Introduction

Data acquisition in single-sink sensor networks might have issues in scalability. As the size of sensor networks grows, the distances between the sink and the responding sensors become larger. This leads to a greater energy consumption for query-flooding and data-collection between sensors and the single sink, leading to a possible reduction in the lifetime of the sensors. Hence, we need to design energy-efficient data acquisition mechanisms that scale with the size of the network.

One solution is to simultaneously deploy multiple sinks in the sensor network. To reduce energy consumption, each sensor would then have an option of sending its data to any one of the sinks by diffusing its data toward its closest sink, as determined by an appropriate energy metric.

In this abstract, we propose a logical-graph model to seamlessly adapt existing single-sink algorithms to multi-sink sensor networks. From the actual multiple-sink sensor graph, we construct a modified graph with a single virtual sink. In particular, we prove that shortest path tree creation in this logical-graph model is equivalent to that in the multi-sink sensor-connectivity graph. We then analytically obtain the expected energy savings for data acquisition for a multi-sink network. We consider both a uniform random sink placement as well as a deterministic grid-based sink placement. While our analysis is restricted to the one and two dimensional cases, we conjecture that, in general, the expected energy savings in a d-dimensional sensor region due to a random placement of k sinks and n sensors is proportional to $k^{\frac{1}{d}}$. We validate our analytical results with the help of simulations.

Very little work done in this area, and to the best of our knowledge, there has been no prior work in the analytical modeling of multi-sink sensor networks. For more details, the reader is encouraged to look at *http://netweb.usc.edu/d̃dutta/research/multi-sink/*.

II. Problem Statement

Given a sensor network with n sensor nodes and k sinks distributed in a field of area A, we pose the following questions:

- Is there a simple mechanism for querying and data-collection in the presence of multiple sinks, without adding extra protocol overheads?
- What are the average energy savings for resolving a query in a multiple-sink sensor network, in comparison with that in a single sink scenario?

We do not assume any in-network aggregation of data. We also assume that the sinks are not resource-constrained. Also, our energy analysis model assumes that the energy needed to transmit a signal over a distance r is proportional to r^2 .

III. Logical Graph Model

In our model, we map a sensor connectivity graph due to a multiple-sink sensor network to a logically equivalent single sink counterpart. For purposes of routing or querying, in this new logical graph, we can view all the sinks as a single logical vertex. For a sensor, this implies that the multiple sinks are indistinguishable from each other, since the routing and connectivity information for the various sinks is advertised as originating from a single virtual sink.

More formally, a multiple-sink sensor network is modeled as an undirected graph $G = (V, E)$ with a subset S of V that are sink-nodes, and edge weights $w(i, j)$ on all edges $(i, j) \in E$. We convert this graph G into another logical graph $G' = (V', E')$ where $V' = V - S + \{s'\}$. s' can be viewed as a single logical vertex representing all the sinks. Thus, we have $E' = \{(u, v)|(u, v) \in E$ and $u, v \notin S\} \cup \{(u, s')|\exists s \in S, (u, s) \in E\}$. We also have a new edge weight function $w'(i, j)$ where $w'(i, j) = w(i, j)$ if $i, j \notin S$ and $w'(i, s') = \min_{\forall s \in S} w(i, s)$ if $j \notin S$

Essentially, we construct a new graph where all the sinks are now considered to be a single logical sink and edges connecting a particular sensor to any sink in G are replaced by an edge from the sensor to the logical sink in G', with its new weight being the minimum cost of the sensor from any sink. The weights here can be chosen appropriately based on the object function to be optimized as a part of the data acquisition (e.g. it could be the communication cost or the error-rate of the link). Now we can adapt various forms of querying or data-collection strategies, that are normally used in the single-sink networks, by using the logical graph G' as its input (instead of G). For example, we can run the *Shortest Path Tree* and the *Minimum Spanning Tree* protocols for the multiple sink case, using existing algorithms that normally work in a single sink scenario. We prove the following theorem and a corollary for adapting a shortest path tree creation algorithm.

Theorem 1: Given a multi-sink sensor graph G, and a non-negative edge-weight function, and the corresponding logical graph G' with a single logical sink s' as described above, a path P, is the shortest path from any sensor m to its closest sink (say s_m) in G if and only if $P' = P - \{s_m\} + \{s'\}$ is the shortest path from the sensor m to the logical sink s' in G'. If w and w' are the weight functions for G and G', then $w(P) = w'(P')$.

Corollary 1: Given the above sensor graph, any distributed Shortest-Path-Tree algorithm that can calculate the shortest path tree from a sensor to the sink in a single-sink sensor network, can also calculate the shortest path tree from a sensor to its closest sink, in the multiple-sink sensor network

Similarly, one can easily adapt other protocols designed for single-sink scenarios, such as distributed minimum spanning tree schemes, minimum broadcast-tree heuristics, maximum-lifetime-routing schemes etc, into the multi-sink framework, by running the corresponding single-sink algorithms on the modified graph G'. Note that our model does not introduce any extra protocol overheads in the process.

IV. Analysis

We now estimate, analytically, the average energy cost savings for a query resolution in the presence of multiple sinks. We assume that every sensor has an equal probability of responding to a given query. The total energy expenditure for a query resolution consists of two parts: the query flooding part, and the data collection part. Since the cost of query flooding is independent of the number of sinks in the network, we therefore only estimate the expected energy expenditure for a query response.

In our analysis, we also assume that the bulk of the sensors' energy consumption for data acquisition is due to their transmission costs, and hence, we use this transmission cost (proportional to d^2, where d is the transmission distance) as a metric when creating routing paths for data acquisition. Additionally, our analysis of the energy savings in this section, is normalized for a unit sensor-to-sink transmission.

A. Random sensor/sink placement

In this abstract, we will only present the results for the random sink and sensor placement. We assume that k sinks are uniformly randomly distributed in a sensor region S of area A. The n sensors themselves are also uniformly randomly distributed in region S. First we consider a one dimensional sensor region and then we move on to the two dimensional case.

1) One-Dimensional Case: We will first consider a one-dimensional line of unit length, where the k sinks as well as the n sensors are uniformly randomly distributed on the line. Given this model, we will estimate the average energy consumption per query response as a function of k. We have the following theorem:

Theorem 2: Given a uniform random distribution of n sensors and k sinks on a straight line, and the sensor communication radius r satisfies the asymptotic connectivity threshold, the average energy costs incurred per query response is asymptotically inversely proportional to $O(k)$

2) Two Dimensional Case: Here, we only consider the two dimensional case, where the n sensors and the k sinks are uniformly randomly distributed within a unit square. We first obtain the expected distance of a random sensor from its closest sink.

Lemma 1: Given a random placement of k sensors and n sinks in a unit square, the expected distance D_{avg} of a random sensor node from its closest sink is given by $D_{\text{avg}} = \frac{1}{1/8}\int_{a=0}^{0.5}\int_{b=a}^{0.5} D(a,b)da.db$, where $D(a,b)$ is given by $\int_{r=0}^{a} 8kr^2(1-4r^2)^{k-1}dr$ $+ \int_{r=a}^{b} kr(4r+2a)(1-2r(a+b))^{k-1}dr$ $+ \int_{r=b}^{1-b} kr(2r+a+b)(1-(r+a)(r+b))^{k-1}dr$ $+ \int_{r=1-b}^{1-a} kr(1-(r+a))^{k-1}dr$. Here, the distance between two locations (x,y) and (a,b) is defined as $D((x,y),(a,b)) = \max(|x-a|,|y-b|)$.

Once we have the expected distance between a random sensor and its closest sink, we now make use of a property proved by Muthukrishnan et. al. in Corollary 3.1 (see our full report), to prove the following lemma:

Lemma 2: Given a random placement of k sensors and n sinks in a unit square region S, the maximum transmission radius r satisfying the asymptotic connectivity threshold, the expected energy consumption , E_{avg}, is proportional to $D_{\text{avg}}.r$, where D_{avg} is as defined before.
While E_{avg} in the above lemma does not have a closed form solution, we can numerically observe that E_{avg} seems to fall off inversely proportional to k^{β}, where k is the number of sinks, and $0.4 < \beta < 0.5$.

V. Simulation Results

We validate our analytical results using a uniformly random sensor and sink placement. First we simulate a one dimensional scenario by placing 200 sensors and a variable number sinks uniformly at random on a unit line. Next we validate our analysis for the two dimension case by considering the same number of sensors and sinks placed uniformly at random on a unit square. We run a Distance Vector protocol in the sensor network using the logical graph model, to find optimal cost paths from sensors to their closest sinks. From the simulations,

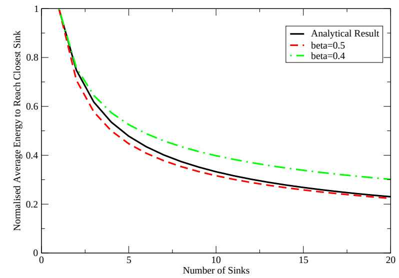
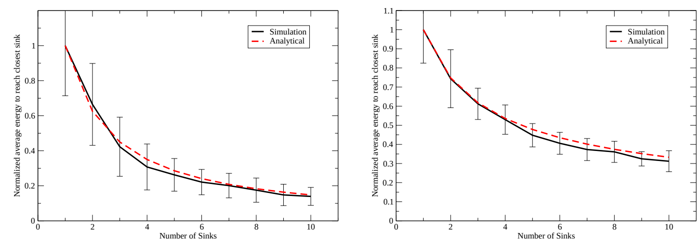


Fig. 1. Plot to compare the normalized average energy for the 2D-case from our analytical model and $k^{-\beta}$ where $\beta = 0.4, 0.5$.



(i) 1D (ii) 2D

Fig. 2. Plot to compare the values from our model and simulations: normalized average energy between a sensor and its nearest sink for (i) 1D and (ii) 2D.

we estimate the normalized communication costs of a single query response, and compare that to our analytical predictions, using our random graph model. We observe in Figure 2 that our analytical model has high accuracy in predicting the expected communication energy consumed in order for a sensor to reach its closest sink.

From our results, two simple observations can be made. First, for 1D, the expected energy spent falls off inversely proportional to the number of sinks, k. Second, in the two dimensional case, the expected energy expenditure seems to fall off inversely proportional to k^{β}, where k is the number of sinks and $0.4 < \beta < 0.5$. Note that k is much closer to 0.5 as shown in Figure 1. We conjecture that the expected energy savings in a d-dimensional sensor region due to deployment of k sinks is proportional to $k^{\frac{1}{d}}$.

VI. Conclusions and Future Work

In this abstract, we have merely scratched the surface of the issues in designing multi-sink sensor networks. As a first step, we have presented a simple logical graph model to adapt algorithms to the multi-sink case, followed by the analytical evaluation as well as the empirical validation of the expected energy savings in such networks for a random placement of sensors and sinks in both one and two dimensions.

There are several interesting directions that include practical protocol design for different optimization functions in multi-sink sensor networks using our logical graph model, and their implementation on a real testbed. Another open area is to consider trade-offs in the presence of in-network data aggregation.

Data Acquisition in Multi-sink Sensor Network

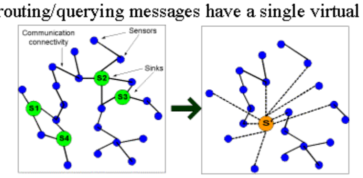
Abhimanyu Das
Debojyoti Dutta

Problem

- Distance between sensors, sink grows with size of the sensor network.
  - So does the communication energy required by energy-starved sensors.
- Need to scale data acquisition systems with size, from an energy efficiency perspective.
  - Increased lifetime for the sensor network.
- One approach: Deploy multiple sinks in the network
  - Sensors send data to closest sinks.
  - Sinks are NOT energy constrained.
- Goals:
  - Design simple data collection models for multiple-sinks to minimize communication costs.
    - __Contribution:__ *New logical graph model. Easy to adapt single-sink algorithms.*
  - What are the avg. energy savings due to multi-sink deployment.
    - __Contribution:__ *Energy costs decrease inversely as* $k^{1/d}$. *Analyzed and validated with simulations for uniformly random placement of sinks and sensors*
- Very little prior work in this area.

Our Approach

- Create a logical sensor graph with a single virtual sink.
  - Contrast to Voronoi tessellation [Estrin04CensTR].
  - Every sensor is aware of the existence of one sink only.
    - All routing/querying messages have a single virtual sink ID



Multiple sink random geometric graph → Logical graph with single virtual sink

$$E' = \{(u,v) \text{ s.t. } (u,v) \in E \text{ and } u,v \notin S\} \cup \{(u,s') \text{ s.t. } \exists s \in S, (u,s) \in E\}$$

$$w'(i,j) = w(i,j) \text{ if } i,j \notin S \text{ and}$$
$$w'(i,s') = \min_{\forall s \in S} w(i,s) \text{ if } j \notin S$$

Protocol Adaptation

- We can adapt protocols designed for single-sinks to multi-sink networks.
  - By running the single-sink algorithm on the virtual graph G`.
  - Eg. Consider Shortest Path Tree creation. We have the following theorem:

Theorem: Given a connected multi-sink sensor graph **G**, and a non-(-ve) edge-weight function **w.** Also, given the corresponding logical graph **G'** with a logical sink **s'** , then
- A path **P**, is the shortest path from any sensor **m** to its closest sink (say $\mathbf{s_m}$) in **G** if and only if **P' = P - $\mathbf{s_m}$ + s** is the shortest path from the sensor **m** to the logical sink **s'** in **G'.**
- **If w and w' are the weight functions for G and G', then w(P) = w'(P').**

- We can therefore use any single-sink shortest path tree protocol on the virtual graph, to calculate shortest distances to closest sinks, in the multi-sink scenario

Analysis: Summary

- Calculated the communication energy saved in contrast to a single-sink sensor network.
- Analytically analyzed simple scenarios.
  - Grid based placement
    - 1D: Savings inversely proportional to **k**
    - 2D: Inversely proportional to $\mathbf{k^{1/2}}$
  - Uniform Random Placement
    - 1D: Savings inversely proportional to **k**
    - 2D: Inversely proportional to $\mathbf{k^d}$, **d~0.5.**
- Used results in Random Geometric Graph **(RGG)**
  - [Muthu03TRdimacs]: If distance between two sensors u,v is ***D(u,v)***, and transmission radius is ***r***, avg. number of hops in the optimal path between u and v = ***2D(u,v,)/r.***
- Validated our analysis using simulations
  - Simulations match analysis for both grid and random graph placement

Sample Analysis: 2-d random geometric sensor network

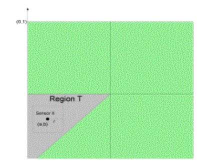
- Lemma: Given a random placement of **k** sensors and **n** sinks in a unit square, the expected distance $\mathbf{D_{avg}}$ of a random sensor node from its closest sink is given by

$$D_{\text{avg}} = \frac{1}{1/8}\int_{a=0}^{0.5}\int_{b=a}^{0.5} D(a,b)\,da.db$$

  - where D(a,b) is given by

$$\begin{aligned} D(a,b) &= \int_{r=0}^{a} 8kr^2(1-4r^2)^{k-1}dr \\ &+ \int_{r=a}^{b} kr(4r+2a)(1-2r(a+b))^{k-1}dr \\ &+ \int_{r=b}^{1-b} kr(2r+a+b)(1-(r+a)(r+b))^{k-1}dr \\ &+ \int_{r=1-b}^{1-a} kr(1-(r+a))^{k-1}dr \end{aligned}$$

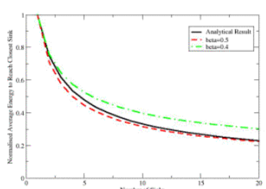
D(a,b) is obtained by considering a single triangular region T of the sensor field, and estimating average distance of a random sensor in that region from its closest sink (using simple probabilistic analysis)



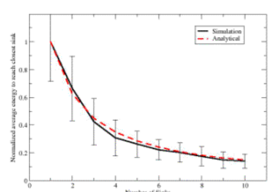
Analytical results

- Theorem: Given a random placement of k sensors and n sinks in a unit square region S, with the sensor communication radius r satisfying the asymptotic connectivity threshold, the expected energy consumption , E_{avg} , in the sensor network for transmitting a query response by a single sensor is proportional to $D_{avg}r$, where D_{avg} is as defined previously
  - uses a stretch property result by Muthukrishnan et al.

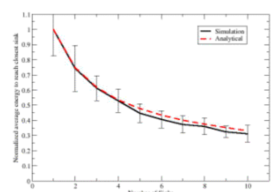
E_{avg} does not have a closed form solution, but numerically can be shown to be inversely proportional to k^d, where 0.4<d<0.5, for realistic values of k (shown in graph)



Simulations



One-dimensional case



Two-dimensional case

- Simulated communication expenditure for query resolution in random geometric graph with 200 sensors and variable number of sinks.
- Ran simple distance vector protocol using our logical-graph model to estimate min-cost paths to closest sink.
- Simulation results match analytical predictions for normalized average communication costs incurred per query resolution, for both one and two dimensional random geometric graph-based sensor networks.

Conclusions and Future Work

- Sensor network based data acquisition systems may not scale with a single sink. Thus, we use multiple sinks.
  - New logical graph model. Easy to adapt single-sink algorithms.
  - Energy savings proportional to $\mathbf{k^{1/d}}$ (**d** is dimension of the sensor region). Analyzed and validated with simulations for uniformly random placement of sinks and sensors
- This work scratched the surface of multi-sink design. Several future directions:
  - Aggregation.
  - Adapting other protocols.
  - Adapting optimization objectives.
  - Real implementation on a testbed.
- More details: http://netweb.usc.edu/~ddutta/research/multi-sink/