Available online at www.sciencedirect.com

ScienceDirect

Ad Hoc Networks 5 (2007) 1317–1328

Ad Hoc Networks

www.elsevier.com/locate/adhoc

Building trees based on aggregation efficiency in sensor networks

Albert F. Harris III [a,*], Robin Kravets [b], Indranil Gupta [b]

[a] *Department of Information Engineering, University of Padova, Italy*
[b] *Department of Computer Science, University of Illinois at Urbana, Champaign, USA*

Available online 15 March 2007

Abstract

Sensor network protocols must minimize energy consumption due to their resource-constrained nature. Large amounts of redundant data are produced by the sensors in such networks; however, sending unnecessary data wastes energy. One common technique used to reduce the amount of data in sensor networks is data aggregation. Therefore, we consider the impact and cost of data aggregation in sensor networks to achieve energy-efficient operation. We propose a new notion of *energy efficiency* that can be used to decide where aggregation points in the network should be placed. The optimal choice of these points is determined by the aggregation efficiency, which determines the amount of data reduction, and the cost in terms of energy to perform the aggregation. We present our aggregation tree algorithm "Oceanus" that produces energy-efficient aggregation trees by taking into account both of these factors. Our evaluation shows that Oceanus provides higher energy efficiency compared to existing solutions.

Keywords: Sensor networks; Aggregation; Routing

1. Introduction

Advances in computing and communication technologies have enabled the creation of small devices capable of complex sensing and computation. While the goal is to embed these devices into our surrounding environments, energy consumption has become the main limiting factor of the lifetimes, and so the effectiveness, of these sensor networks. To support increased network lifetime, it is necessary to design energy-efficient communication protocols. Although such protocols have been proposed in the context of ad hoc networks [1], the data-centric focus of sensor networks lends itself to better energy efficiency through intelligent management of the data.

In typical communications scenarios for sensor networks, data about a particular event is collected by the sensors and is then sent to a data sink, which can be anywhere in the network. Frequently, the sink may not require the original data from each individual sensor, but instead only require an aggregate function (e.g., sum, average, etc.) of the collected data from all sensors. The benefit of such *data aggregation* is that it can reduce the total amount of data sent through the network, increasing network performance and decreasing energy consumption. However, the overall effectiveness of data aggregation is dependent on where and when the aggregation actually occurs.

* Corresponding author.
E-mail addresses: harris@dei.unipd.it (A.F. Harris III), rhk@cs.uiuc.edu (R. Kravets), indy@cs.uiuc.edu (I. Gupta).

1570-8705/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2007.02.021

Data aggregation changes the communication in the network by allowing individual nodes to collect data samples from multiple sources and combine them to be transmitted as one sample. Energy can be saved if the overall amount of data transmitted in the network is reduced by the aggregation. Therefore, the *energy efficiency* of such aggregation is affected by two metrics *aggregation efficiency* and *aggregation cost*. Aggregation efficiency captures the amount of data compression achieved by the aggregation function. If the aggregation of n data samples results in one new data sample, the aggregation efficiency is said to be perfect. However, if the result is simply the n samples concatenated together, the aggregation efficiency is poor, and only benefits from merging headers. Although aggregation may be highly efficient and so significantly reduce the amount of data transmitted, it is also necessary to consider the computational cost of the aggregation in the node. While some aggregation may be cheap (e.g., simple sum), some aggregation may be computational expensive (e.g., combining audio samples).

Since the goal of data aggregation is to reduce redundancy in the communication, the best-suited delivery network is a tree, where aggregation occurs when two branches merge. The challenge, therefore, is to design algorithms that take into account aggregation efficiency and cost to create trees with the most energy-efficient aggregation points. As discussed in Section 5.1, if the aggregation algorithm is perfect (*i.e.*, perfect efficiency and 0 cost), the optimal aggregation tree is a Steiner Tree. Given an imperfect aggregation algorithm (*i.e.*, less efficient and some cost), the optimal aggregation tree is a Weighted Steiner Tree. Although calculating a Steiner Tree, weighted or unweighted, is NP-complete [2], it is possible to use heuristics to approximate the target Steiner Tree and use this approximation as the aggregation tree in the sensor network.

To find the most energy-efficient aggregation tree, it is necessary to understand the energy efficiency of the data aggregation algorithm. In this paper, we use our formulation of aggregation efficiency and cost to explore the energy efficiency of data aggregation. Essentially, we show that if the aggregation efficiency is perfect and the cost is free, the aggregation points should be as close as possible to the sources to save the most energy. However, as the aggregation efficiency degrades or the cost increases, the optimal aggregation points drift towards the sink, since the savings from the reduced communication no longer outweighs the extra cost of aggregation.

The main contributions of our research are a mathematical analysis of when aggregation should be performed and the design and analysis of Oceanus [3], a heuristic-based aggregation tree algorithm that approximates the optimal Weighted Steiner Tree for a given aggregation efficiency and cost. By understanding the tradeoffs between aggregation efficiency and cost, Oceanus creates trees with aggregation points closer to the sources when efficiency is high and cost is low and trees with aggregation points closer to the sink when efficiency is low and/or cost is high. Our evaluation of Oceanus shows that for most aggregation scenarios, Oceanus saves energy over the shortest path tree, the opportunistic method, and the greedy incremental method. We also notice that in extreme cases, where the sources are topological isolated from each other, the opportunistic method outperforms all others, since little or no aggregation is possible. This is a limitation of the heuristic-based algorithm. However, in such cases, if the disjoint sources are treated independently, Oceanus can again outperform the other approaches.

The remainder of this paper is structured as follows. Section 2 presents a brief overview of related work. Section 3 presents a mathematical model of in-network data aggregation. Section 4 uses the model to answer the question "When should data aggregation be performed?" Section 5 presents four aggregation tree algorithms: the optimal tree algorithm and the three heuristic algorithms, including Oceanus. Section 6 presents the methodology used to analyze the aggregation efficiency space as well as the simulation setup and our experimental results. Finally, Section 7 presents conclusions and future directions for our research.

2. Related work

Although several data aggregation algorithms and frameworks have been proposed [4–8], finding the optimal aggregation points in the network is still an open area of research.

Current heuristic-based aggregation tree algorithms use either opportunistic methods (e.g., Directed diffusion [9]) or greedy incremental methods (e.g. Intanagonwiwat et al. [10,11]). In opportunistic methods, data flows through shortest paths from the sources to the sink. In the event that paths meet, the paths are joined to form an aggregation point. Such aggregation points tend to be close to the sink because shortest path flows from different sources to the same sink intersect downstream. In the greedy incremental methods, one source initiates a shortest path flow to the sink. Then, the other sources connect to that path via shortest paths, which generally results in aggregation points closer to the sources. However, the efficiency of the greedy incremental method is entirely determined by the first path and can result in very inefficient aggregation trees. The main problem with both of these methods is that they cannot consider aggregation cost and efficiency.

Many algorithms have been proposed to enable efficient interest notification ([4,5,7,8,12–14]) as well. These are a critical component to any sensor network scheme but largely orthogonal to the work in this paper.

Zhu et al. [15] present an analysis of the limits on achievable energy improvements through the use of data aggregation. They use Steiner trees, as in this work, as the optimal base for their analysis.

Finally, Snader et al. [16] present a distributed algorithm intended to locate the optimal points of aggregation in the network. While their algorithm performs better than opportunistic methods, it performs significantly worse than optimal.

3. Energy-efficiency and data aggregation

Energy efficiency is a driving concern in the design and implementation of sensor networks. When using data aggregation, there are three components to energy consumption in sensor networks: the energy consumed by control messages to set up the aggregation tree for a given event, the energy consumed by data transmission and reception for a given event, and the energy consumed by the aggregation of the data at the aggregation points. While the energy consumed by the control messages is relatively fixed for a given network, there is a direct tradeoff between the energy consumed by the data transmissions and the energy consumed by the aggregation. In this section, we define the energy efficiency of data aggregation, which captures this tradeoff. In the following section, we show how our model can be used to determine when aggregation should be performed.

For a given event, data must flow from the source to the sink, as determined by an initial interest notification sent from the sink. For the non-opportunistic approaches, some energy is consumed during coordination between the sources and the sink to set up the aggregation tree. Since sensor networks are relatively static, this paper focuses on the energy efficiency of static aggregation tree algorithms. If enough nodes die so that the current tree can no longer deliver the data, it is necessary to reconfigure the tree, incurring some energy consumption from control messages. We are currently investigating the impact of this control overhead and developing a dynamic aggregation tree algorithm.

Let $G = (V, E)$ be a graph. Let the sources in the network be $s_i \in S \subseteq V$ and the sink be $k \in V$. Then the graph M with the minimum cost in terms of energy can be identified by adding a weight w to each edge ($e_{i,j} \in E$). Let, $w_{i,j}$ be composed of the cost to receive the data at node i plus the cost to aggregate the data plus the cost to transmit the data to node j. The acyclic path from all s_i to k such that $\sum_{\forall i,j} w_{i,j}$ is minimized will be the optimal aggregation tree in terms of energy efficiency.

Once sinks and sources have been coupled in the network, data can flow from the sources to the sink(s). The energy consumed at each hop of a flow is determined by the transmission rate (R), the size of the data to be transmitted (x), where x includes a header (H) and the data (D), and the power consumption of the network interface while transmitting (P_t). Therefore, the transmission energy to send data from node i to node j is defined as follows:

$$T_{i,j}(x) = \frac{x}{R} \times P_t. \tag{1}$$

Whenever flows from the same event intersect, it is possible to combine their data into one flow. The resulting flow carries the aggregated data (D_a) and only uses one header (H). For example, if n flows are being aggregated, where a packet from flow m is $[H_m, D_m]$, the packets from the aggregated flow are $[H, D_a]$, where $D_a = f(D_1, D_2, \ldots, D_n)$.

Although the goal is to reduce the number of bytes sent and so reduce the energy consumed by the transmission, it is possible that $T_{i,j}$ could increase after the aggregation if

$$(H_a + D_a) > \sum_{m=1}^{n}(H_m + D_m). \tag{2}$$

In other words, if the aggregation function results in an aggregate that is larger than the sum of the initial data sizes plus the size of $n - 1$ headers, the aggregation function should not be used.

To measure the effectiveness of the aggregation function, we define aggregation efficiency as follows:

$$\delta(x,y) = \frac{x_{\text{in}}}{x_{\text{out}}}, \tag{3}$$

where x and y are the sizes of the two packets of data to be aggregated, including headers, $x_{\text{in}} = x + y$, and x_{out} is the total size after aggregation including the header. If $\delta(x,y) > 1$, the aggregation reduces the number of bytes transmitted, and so reduces $T_{i,j}$. If $\delta(x,y) < 1$, aggregation should not be performed.

While $T_{i,j}$ captures the energy to transmit, it is also necessary to consider the cost of receiving the data and the cost of executing the aggregation function. The cost of receiving the data is also a function of the data size (x) and the power use to receive (P_r), and is given by:

$$R_{i,j}(x) = \frac{x}{R} \times P_r. \tag{4}$$

Therefore, reducing the data size via data aggregation will also reduce the cost of receiving the data by the downstream nodes. The cost of running the aggregation function at node i ($c_i = C(x,y)$) must be low enough so that it does not outweigh the savings in $T_{i,j}$ and $R_{i,j}$ from the aggregation. The cost function ($C(x,y)$) depends on the power consumption of the processor, the frequency of the processor, the number of cycles required by the aggregation function, etc.

To analyze the effects of aggregation, the resulting data size and aggregation costs for a group of packets must be modeled. Let x_m be the sizes of packets in node i's queue and n be the total number of packets in the queue. If no aggregation is performed then total size of data to be sent is simply:

$$x = \sum_{m=1}^{n} x_m, \tag{5}$$

and the cost of aggregation (c_i) is 0.

However, if aggregation is used, then the aggregate size of the n packets (a_n) is given recursively, where $a_1 = x_1$ and x_n is the size of the nth packet, by:

$$a_n = \frac{(a_{n-1} + x_n)}{\delta(a_{n-1}, x_n)}. \tag{6}$$

The cost to aggregate at node i (c_i), where $a_1 = x_1$, is

$$c_i = \sum_{m=2}^{n} C(a_{m-1}, x_m), \tag{7}$$

and the single link cost is defined by:

$$c_L = c_i + T_{i,j}(x_a) + R_{i,j}(x_a). \tag{8}$$

The energy efficiency (ξ) of a data aggregation algorithm performed on n packets of data can then be defined as follows:

$$\xi = \frac{T_{i,j}\left(\sum_{m=1}^{n} x_m\right) + R_{i,j}\left(\sum_{m=1}^{n} x_m\right)}{c_i + T_{i,j}(a_n) + R_{i,j}(a_n)}. \tag{9}$$

With these definitions in hand, the question of when aggregation should be performed can be answered for various cases.

4. Should aggregation be performed?

The question of whether or not aggregation should be performed can be broken down into two cases. The first case is if aggregation can be performed without having to re-route data (*i.e.*, opportunistically) and the second case is if aggregation can be performed only if the data is re-routed to aggregation points. This section handles each of these cases in turn. For each case, assume that each individual, un-aggregated flow begins by traveling to the sink via a shortest path. Furthermore, assume that to transmit and receive an amount of data x between any two hops is the same.

4.1. Case 1: No re-routing

The first case involves no changes in path lengths, and therefore simply comes down to whether or not the savings in sending and receiving data outweigh the energy costs of performing the data aggregation. Let x_m be the packets in node i's queue. Let h be the number of hops from the aggregation point to the sink, then if

$$\left[h\left(T_{i,j}\left(\sum_{m=1}^{n} x_m\right) + R_{i,j}\left(\sum_{m=1}^{n} x_m\right)\right) - h(T_{i,j}(a_n) + R_{i,j}(a_n))\right] > C_i(\forall m, x_m),$$

where a_n is given by Eq. (6), then aggregation should be performed.

4.2. Case 2: Re-routing is required

The second case can be further divided into two sub-cases: one, re-routing maintains shortest path lengths and two, re-routing increases the path lengths.

4.2.1. Re-routing maintains path lengths

Let the number of hops between a source (s) and the sink (k) be h. Then the transmission energy from

s to k ($T_{s,k}$) assuming no use of transmit power control is given by:

$$T_{s,k}(x) = hT_{s,j}(x), \tag{10}$$

where node j is the first hop away. The total transmission energy for n flows (s_m), each h_m hops away respectively without aggregation is as follows:

$$T_{s_n,k}(x_n) = \sum_{m=1}^{n} h_m T_{s_m,j}(x_m), \tag{11}$$

and the total cost of aggregation is trivially 0. Now for aggregation, assume an aggregation point is located at node g, which is $h_{s_m,g}$ hops from each of the s_m sources and $h_{g,k}$ hops away from the sink. Then the total transmission energy $T_{s_n,k}(a_n)$ is given as follows:

$$T_{s_n,k}(a_n) = \left[\sum_{m=1}^{n} h_{s_m,g} T_{s_m,j}(x_m)\right] + [h_{g,k} T_{g,j}(a_n)], \tag{12}$$

where a_n is given by Eq. (6).

But, since none of the path lengths have changed, the energy savings (E_s) can be given by the savings over the hops for which the data was aggregated as follows:

$$E_s = \left[\sum_{m=1}^{n} h_{g,k}(T_{s_m,j}(x_m) + R_{s_m,j}(x_m))\right] - [h_{g,k}(T_{g,j}(a_n) + R_{g,j}(a_n))], \tag{13}$$

and aggregation should be performed if

$$E_s > C_g(\forall m, x_m). \tag{14}$$

4.2.2. Re-routing increases path lengths

Again, let the number of hops between a source (s) and the sink (k) be h. Then the transmission energy from s to k ($T_{s,k}$) is given as in Eq. (11). However, the case for aggregation is more complex, essentially there is an extra term that must be considered in Eq. (13). For each flow, any increase in path length must be taken into account, therefore, let n'_m be the new path lengths. Then the energy consumed due to the extra path lengths can be expressed as

$$\sum_{m=1}^{n} (h'_{s_m,g} - h_{s_m,g})(T_{s_m,j}(x_m) + R_{s_m,j}(x_m)). \tag{15}$$

Notice that this equation takes into account the possibility that some paths may also be shortened. The total energy savings in the face of path length changes is then:

$$E_s = \left[\sum_{m=1}^{n} h_{g,k}(T_{s_m,j}(x_m) + R_{s_m,j}(x_m)) - h_{g,k}(T_{g,j}(a_n) + R_{g,j}(a_n))\right] + \left[\sum_{m=1}^{n} (h'_{s_m,g} - h_{s_m,g})(T_{s_m,j}(x_m) + R_{s_m,j}(x_m))\right], \tag{16}$$

and aggregation should be performed if

$$E_s > C_g(\forall m, x_m). \tag{17}$$

Clearly, the choice to aggregate depends on the relationship between the aggregation efficiency, the cost of aggregation, and the path length increases. This is the most complex case and choosing optimal aggregation trees in the face of this constraint is NP-complete, which will be discussed in Section 5.1, therefore, even a centralized solution must use some heuristic methods.

Essentially, as the aggregation efficiency increases and the cost of the aggregation decreases, it is more efficient to aggregate close to the sources. As the aggregation efficiency decreases and the aggregation cost increases, the most efficient aggregation points migrate towards the sink. Current research, however, tends to look at algorithms for creating aggregation trees and aggregation energy efficiency in isolation. In the next section, we present various aggregation tree algorithms and discuss if and how they can integrate energy efficiency.

5. Aggregation tree algorithms

The focus of this work is on the effects of data aggregation efficiency on choosing the aggregation tree in sensor networks. There has been significant work in the areas of finding clusters of nodes among which to shift aggregation points to increase network lifetime [17–20] and building general routing policy frameworks [6], but these are orthogonal to this work. In this section, we describe the optimal aggregation tree algorithm and then consider three heuristics used to construct aggregation trees in sensor networks. The three heuristics are opportunistic, greedy incremental, and Oceanus. In the following sections, we evaluate the effect of energy efficiency on each of these algorithms.

5.1. Optimal aggregation

The problem of finding a lowest cost aggregation tree can trivially be reduced to the problem of finding a Steiner tree in the graph. Formally, the Steiner tree of some subset of the vertices of a graph G is a minimum-weight connected subgraph of G that includes all of the vertices.

Let $G(V, E)$ be the set of all nodes in the sensor network with non-negative weights for each $e \in E$ corresponding to the cost to transmit data over link e. Let $Z \subseteq V$ be the set of source nodes ($s_i \in S$, where $S \subset Z$) and the sink (k). The cost function is in terms of energy consumed on each link for transmission of data across that link. If the cost function for assigning the weights includes data aggregation (see Sections 3 and 4), finding the subnetwork $T \subseteq G$ such that every pair of vertices in Z is connected and the total cost of T is a minimum is equivalent to finding the minimum cost aggregation tree. However, this is the Weighted Steiner Tree.

The determination of a Weighted Steiner Tree is NP-complete [2]. Even if the edge weights are all equal (corresponding to a perfect aggregation algorithm), the problem remains NP-complete. Therefore, it is infeasible to calculate the optimal aggregation tree and heuristics for efficiently constructing aggregation trees are needed.

5.2. Opportunistic aggregation

Opportunistic aggregation only aggregates streams if they happen to intersect on their way to the source, (e.g., Directed diffusion [9]). To achieve opportunistic aggregation, each source begins sending streams to the receiver via shortest path routes. As streams intersect, they are aggregated.

Aggregation points are always downstream, towards the sink. The more dispersed the source nodes are from each other in the network, the less likely aggregation is performed. Such trees are beneficial if the energy efficiency of the aggregation is low, meaning that the savings from aggregating early does not outweigh the additional communication cost. However, opportunistic aggregation is less likely to result in energy efficient aggregation trees for aggregation functions with high efficiency and low cost.

5.3. Greedy incremental aggregation

The Greedy incremental aggregation algorithm (e.g., Intanagonwiwat et al. [10]) begins by sending a single stream via a shortest path route to the sink. Each additional stream is then routed to a node participating in the first flow via a shortest path. This method prevents two streams from spanning the entire network only one hop apart.

Aggregation trees derived from this sort of algorithm often have aggregation points that lie somewhere in the middle of the network. This can yield significant efficiency gains over opportunistically created aggregation trees in cases where the aggregation algorithm has moderate to high energy efficiency.

5.4. Oceanus

Oceanus approximates the aggregation tree providing the most energy-efficient communication using knowledge of the energy efficiency of the aggregation algorithm. Oceanus uses a heuristic-based algorithm that approximates a Weighted Steiner Tree, where the weights reflect the energy efficiency of the aggregation algorithm. To start, Oceanus randomly chooses one of the source nodes. It then finds the node in Z that is closest to the chosen node using a shortest path algorithm where the weights on the paths are given by the energy model in Section 3. These nodes are connected by this path. Then, the next node in Z that is closest to the tree that has been already formed is chosen and connected, and so on until a complete tree is obtained. The aggregation tree is calculated according to the algorithm depicted in Fig. 1.

It is simple to see that this method results in each closest sensor node being connected via a least-cost path, where cost is defined in terms of Eq. (8). Since this is only a heuristic, it is possible that this is not the Weighted Steiner tree. If two nodes can be connected via two different least-cost paths, the intermediate node that is chosen may be farther away from all of the remaining unconnected nodes than the other choice of intermediate node.

To find the aggregation tree, we assume that the sink node has knowledge of the sensor network topology. When an event of interest happens, each sensor node sends an initial notification of event to the sink node. At that time, the sink calculates the aggregation tree and informs the source nodes and the aggregation nodes of the paths to follow. Currently, an aggregation tree is calculated for each independent event in the network, for each sink. This is an implementation detail however, and could be altered in future versions.

```
Calculate-Tree()
    $G$:Set of Nodes
    $Z \subseteq G$: Source Nodes + Sink Node
    $S \subseteq Z$: Source Nodes
    $T$: Nodes in Tree
    $T = \{\}$
    $T' \subseteq Z$: Source and Sink nodes in Tree
    $T' = \{\}$
    $z \in S$: $z$ is chosen randomly
    $x \in Z$: $x$ is closest to $z$
    Connect($x$,$z$)
    $T = T + x + z +$ Path($x$,$z$)
    $T' = T' + z + x$
    while($T' \neq Z$)
        $x \in Z$: $z$ is closest to $z \in T$
        Connect($x$,$z$)
        $T = T + x +$ Path($x$,$z$)
        $T' = T' + x$
```

Fig. 1. Oceanus aggregation tree algorithm.

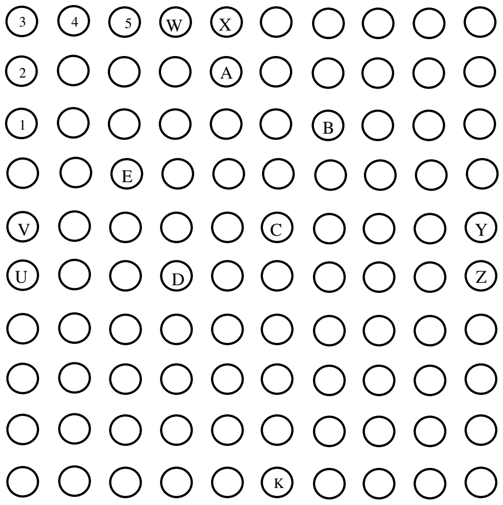

Fig. 2. Network event scenarios (Scenario 1: nodes 1–5 are sources, Scenario 2: nodes A–E are sources, Scenario 3: nodes U–Z are sources).

6. Evaluation

In this section, we analyze the three data aggregation tree algorithms. The goal of this analysis is to determine the most energy-efficient aggregation tree given varying levels of aggregation efficiency.

We analyze the aggregation tree algorithms in terms of the amount of energy spent transmitting sensor data. Because the algorithms are implemented in the ns2 network simulator [21], the link energy model in Eq. (8) is used to provide the energy analysis.

6.1. Simulation setup

The sensor network consists of a 100 node network laid out on a grid. Each node in the middle of the grid has 8 one-hop neighbors. The sensor data size is 64 bytes and the packet header size is 6 bytes. Each node has a transmit power of 36 mW and a receive power of 5.4 mW. These values were chosen to model a common sensor node [22]. The data transmission rate of the nodes is 40 Kbps.

We consider three sensor scenarios. The first scenario is where an event occurs at the corner of the sensor network (see Fig. 2, nodes 1–5 are sources). This is a common model for data aggregation studies. The second scenario is where an event occurs in the middle of the network (see Fig. 2, nodes A–E are sources). The final scenario is where events occur at the edges of the network (see Fig. 2, nodes U–Z are sources).

In all scenarios, the sink is placed near the bottom edge of the network. For the simulations, no mobility is used since we assume a static sensor network. Additionally, we do not consider node failure. Finally, all results are the averages of 30 runs, with the sink nodes being placed randomly along the edge of the network and the source nodes being placed in different spots throughout the network, but in the same configuration, depending on the scenario. In this work, we do not present results varying the number of source nodes, however, these results are similar until the number of source nodes approaches 20% of the nodes in the network.

6.2. Experimental results

Two groups of results are presented in this section. The first group evaluates the performance of Oceanus, the opportunistic algorithm, the greedy incremental algorithm, and sending the data via shortest path routes for a perfect aggregation algorithm with no cost. The second group of results evaluates the performance of these methods across aggregation algorithms with varying aggregation efficiencies, but no aggregation cost. Varying the aggregation cost has the same effect. For all experiments presented here, we consider only the energy expenditure of nodes within the aggregation tree.

This is reasonable because the focus of this paper is finding the minimum cost aggregation trees in terms of energy. Developing algorithms for putting nodes to sleep that are not part of the aggregation tree is outside the scope of this work. A 95% confidence interval of ±2% was maintained for all results.

6.2.1. Perfect aggregation

The results in this section use a perfect aggregation function. There is no cost for aggregation and the amount of data sent after aggregation is 64 bytes, the same as a unit of sensor data, no matter how many flows are being aggregated. Therefore, the optimal aggregation trees aggregate flows as close to the sources as possible. These experiments explore the most efficient end of the aggregation spectrum. For each of the three algorithms tested, opportunistic, greedy incremental, and Oceanus, the algorithms are run over 10 varying networks conforming to the three basic configurations. The results presented are the average results from these runs. The graphs in this section are normalized to the energy expenditure in mJ of sending each sensor's data via a shortest path link with no aggregation. This yields percentage savings over the shortest path, no aggregation method for each of the algorithms.

Fig. 3 depicts the energy savings for networks where the sensors are clustered in a corner of the network. Because the nodes are clustered, the opportunistic algorithm has a reasonable likelihood of causing two streams to flow to the sink only one hop away from each other. However, the greedy incremental algorithm is able to aggregate the flows somewhat near to the sources. Oceanus, on the other hand, links all of the nodes together in a chain, aggregating at the sources, and transmits through the closest node to the sink with high probability. Oceanus uses 26% less the energy then the baseline and about 15% less than the greedy incremental algorithm.

Fig. 4 depicts the energy savings for networks where the sensors are surrounding an event in the middle of the network. In this scenario, it is likely that the nodes on the side of the event away from the sink will aggregate with nodes closest to the sink rather early. For the greedy incremental algorithm, it is more likely that the nodes aggregate close to the sources. Oceanus sends data around the event in a ring, aggregating at each source, and then connects the ring to the sink via one of the source nodes. Oceanus uses 40% less energy then the baseline and about 10% less energy than the greedy incremental algorithm.

Fig. 5 depicts the energy savings for networks where the sensors are on different edges of the network. This scenario represents a difficult case. With high probability, the opportunistic algorithm can only aggregate those nodes in each sector of the network. Therefore, it is likely that multiple independent streams will be sent through the network without aggregation. The greedy incremental algorithm will likely send all data across the center of the network, aggregating in the middle, depending on which stream begins first. Oceanus routes data around the edges of the network, in a large circle, from source node to source node, aggregating at the sources. Oceanus uses 43% less energy than the

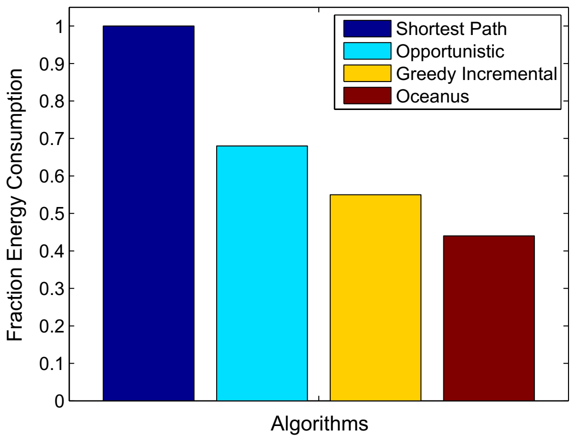

Fig. 3. Perfect aggregation, Network 1.

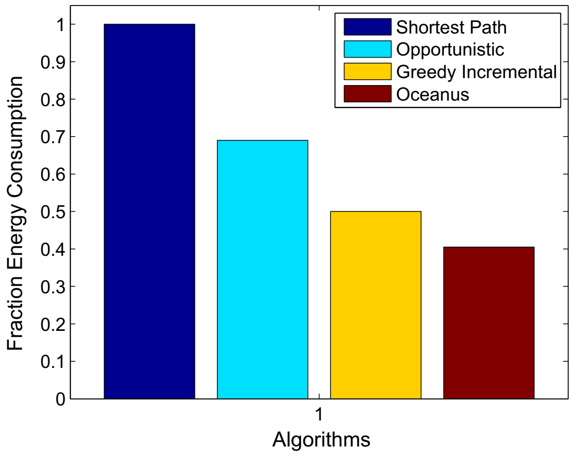

Fig. 4. Perfect aggregation, Network 2.

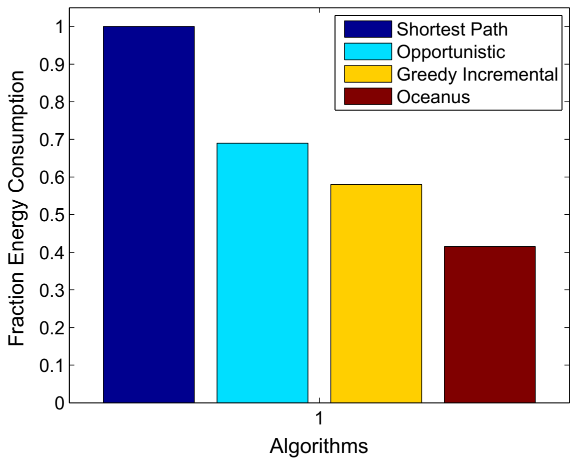

Fig. 5. Perfect aggregation, Network 3.

baseline and about 10% less energy than the greedy incremental algorithm. This scenario is the worst case for the opportunistic algorithm. As expected, the greedy incremental method and Oceanus perform well, with Oceanus providing greater energy efficiency by routing through all of the source nodes. Therefore, for perfect aggregation functions, Oceanus outperforms all other aggregation methods.

The total energy consumed for varying lengths of data flows from 1 sensor data packet from each sensor to 100 packets from each sensor was also tested for perfect aggregation. The longer the flows go on, the cost of setting up the aggregation tree is amortized over the length of the flow. For each of the three algorithms tested, opportunistic, greedy incremental, and Oceanus, we ran the algorithms over 10 varying networks conforming to the three basic configurations. This showed the trends in the data over a range of lengths of flows. Since none of the aggregation trees can be set up before the first packets are received, all of the algorithms have the same energy consumption as the Shortest Path flows. However, as the flows progress, each of the algorithms' energy consumptions diverge. As expected, the results remained the same as presented in Figs. 3–5, with the gaps between the tree algorithms increasing roughly linearly as the flow lengths increased.

6.2.2. *Varying the aggregation efficiency*

The locations of the optimal aggregation points depends on the efficiency of the aggregation algorithm. Oceanus attained the most efficient communication for perfect aggregation functions by aggregating data close to the source nodes. However, as the efficiency of the aggregation algorithm decreases, the optimal aggregation points move closer to the sink node.

The graphs in this section present the amount of energy to send 100 sensor packets from each sensor node to the sink using the three aggregation methods (opportunistic, greedy incremental, and Oceanus) as well as shortest path routing. The *x*-axis of the graphs represent the amount of data compression achieved by the aggregation function. One represents a perfect aggregation function and zero represents simple concatenation.

Fig. 6 depicts the energy consumed for varying aggregation efficiencies in a network where the sources are in the corner of the network. The opportunistic algorithm curve is rather linear, as expected. This is because the data aggregation is performed only if shortest paths from the source nodes cross. Therefore, it is affected least by changes in the aggregation efficiency. Both the greedy incremental method and Oceanus begin to perform more poorly as the aggregation efficiency decreases. This is because the benefit of aggregation shrinks below the possible increase in hops a flow makes to reach an aggregation point. However, because the sources are collected in a corner of the network, no path is lengthened significantly. Therefore, both the greedy incremental algorithm and Oceanus continue to outperform the opportunistic by a significant margin, with Oceanus always being the most efficient.

Fig. 7 depicts the energy consumed for varying aggregation efficiencies in a network where the

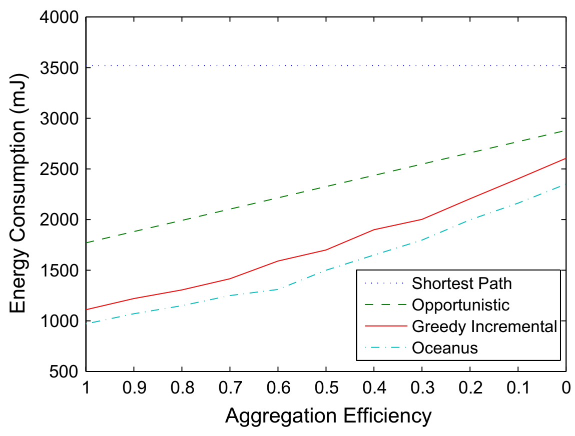

Fig. 6. Energy consumption *vs.* aggregation efficiency, Network 1.

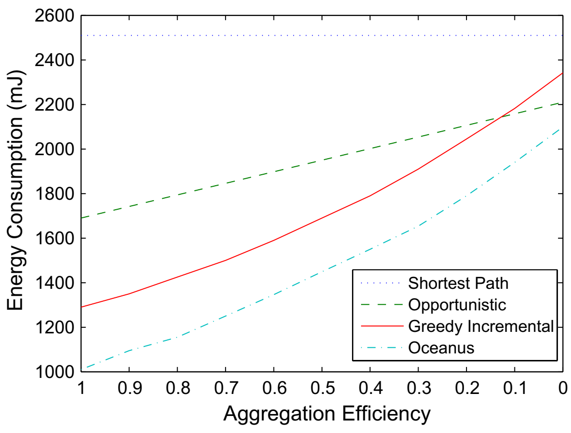

Fig. 7. Energy consumption *vs.* aggregation efficiency, Network 2.

sources are circled around an event in the middle of the network. Again, the curve for the opportunistic method is roughly linear as expected and both the greedy incremental algorithm and Oceanus begin to suffer as the aggregation algorithm becomes less efficient. This time however, the greedy incremental algorithm begins to consume more energy than the opportunistic algorithm for efficiencies close to simple concatenation. This is because some paths are lengthened to reach aggregation points, but the gain in aggregation no longer outweighs these path increases. Oceanus continues to be the most efficient algorithm in this case as well. This is because its gains in aggregation follow along the shortest links to the source node and therefore still outweigh any increases in path length.

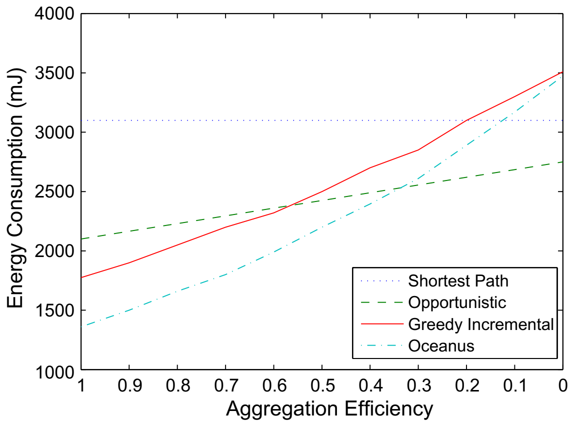

Fig. 8. Energy consumption *vs.* aggregation efficiency, Network 3.

Fig. 8 depicts the energy consumed for varying aggregation efficiencies in a network where the sources are around the sides of the network. As the graph shows, it is the worst case for Oceanus at poor aggregation efficiencies. Around 82% aggregation efficiency, the opportunistic algorithm begins to outperform Oceanus. This is because, the Oceanus algorithm always tries to perform some aggregation, but in this case, this causes poor performance for the worst aggregation efficiencies. However, Oceanus continuously outperforms the greedy incremental algorithm.

6.2.3. Summary

Oceanus significantly outperforms both the opportunistic method as well as the greedy incremental method of aggregation for perfect aggregation functions. Furthermore, in all three network scenarios, Oceanus outperforms the greedy incremental method for all aggregation function efficiencies. However, for aggregation efficiencies in the worst 18%, the opportunistic method outperforms Oceanus in the scenarios where the sources are scattered on all sides of the network. However, this scenario is unlikely, since it is rare that data from events occurring in completely different areas of the network will be aggregated. Therefore, in most realistic scenarios, Oceanus outperforms both the opportunistic method and the greedy incremental method of data aggregation. This is because Oceanus takes into account the energy efficiency of the aggregation algorithm in creating its aggregation trees. Finding efficient data aggregation algorithms is itself the topic of other work; however, the more efficient the aggregation algorithms, the more efficiently data can be transmitted through a sensor network.

One important point to bring up is that Oceanus often times aggressively aggregates data in the network. This means that the loss of an aggregate packet results in the loss of a great amount of data. There may be a desire to trade off the amount of loss with the energy savings from aggregation efficiency. To this end, algorithms to decide what data is aggregatable could be used to partition data of the same types into independent groups, essentially forcing redundancy in the network. The creation of such algorithms is beyond the scope of this paper and is also orthogonal to the Oceanus algorithm.

7. Conclusions and future work

This paper has explored the aggregation efficiency space and where aggregation points should be located within a sensor network to provide energy efficient communication. We have demonstrated that for high energy efficient aggregation algorithms, aggregation points should lie close to the sources of the data. However, as aggregation efficiency decreases, aggregation points should be migrated towards the sink. Therefore, we present Oceanus, which builds the aggregation trees based on the efficiency of the aggregation algorithm. We have shown that Oceanus outperforms both the opportunistic and the greedy incremental methods over a range of network topologies and aggregation efficiencies.

This analysis suggests a future direction for aggregation tree algorithm design. A distributed algorithm that migrates the aggregation points towards the source for high-efficiency aggregation algorithms and towards the sink for low efficiency algorithms is a future direction. Future work also consists of finding energy efficient means of taking care of failures in the network and adding load-balancing. We have not implemented any load balancing mechanism, or fault tolerance into Oceanus.

References

[1] C. Sengul, R. Kravets, Conserving energy with on-demand topology management, in: Second IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 2005.
[2] M. Garey, D. Johnson, Computers and Intractability, Freeman, San Francisco, CA, 1979.
[3] A. Harris, R. Kravets, I. Gupta, Building tress based on aggregation efficiency in sensor networks, in: The IFIP Fifth Annual Mediterranean Ad Hoc Networking Workshop (Med Hoc Net), 2006.
[4] P. Bonnet, J. Gehrke, P. Seshadri, Towards sensor database systems, in: Mobile Data Management, 2001.
[5] S. Madden, M. Franklin, J. Hellerstein, W. Hong, Tag: A tiny aggregation service for ad-hoc sensor networks, in: OSDI, 2002.
[6] R. Kumar, M. Wolenetz, B. Agarwalla, J. Shin, P. Hutto, A. Paul, U. Ramachandran, Dfuse: a framework for distributed data fusion, in: Proceeding of the First International Conference on Embedded Network Sensor Systems, 2003.
[7] Y. Yao, J. Gehrke, The cougar approach to in-network query processing in sensor networks, in: SIGMOD, 2002.
[8] S. Madden, M. Franklin, J. Hellerstein, W. Hong, The design foy an acquisitional query processor for sensor networks, in: SIGMOD, 2003.
[9] C. Intanagonwiwat, R. Govindan, D. Estrin, Directed diffusion: a scalable and robust communication paradigm for sensor networks, in: MobiCom, 2000.
[10] C. Intanagonwiwat, D. Estrin, R. Govindan, J. Heidemann, Impact of network density on data aggregation in wireless sensor networks, in: ICDCS, 2002.
[11] B. Krishnamachari, D. Estrin, S. Wicker, The impact of data aggregation in wireless sensor networks, in: DEBS, 2002.
[12] M. Chu, H. Haussecker, F. Zhao, Scalable information-driven sensor querying and routing for ad-hoc heterogeneous sensor networks, in: Intl. J. High Performance Computing Applications, 2002.
[13] R. Govindan, J. Hellerstein, W. Hong, S. Madden, M. Franklin, S. Shenker, The sensor network as a database, Tech. Rep. 02-771, Computer Science Department, University of Southern California, (September) 2002.
[14] N. Sadagopan, B. Krishnamachari, A. Helmy, The acquire mechanism for efficient querying in sensor networks, in: SNPA, 2003.
[15] Y. Zhu, K. Sundaresan, R. Sivakumar, Practical limits on achievable energy improvements and useable delay tolerance in correlation aware data gathering in wireless sensor networks, in: IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2005.
[16] R. Snader, A. Harris, R. Kravets, Tethys: a distributed algorithm for intelligent aggregation in sensor networks, in: IEEE Wireless Communications and Networking Conference (WCNC), 2007.
[17] Q. Fang, F. Zhao, L. Guibas, Lightweight sensing and communication protocols for target enumeration and aggregation, in: MobiHoc, 2003.
[18] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: Proceedings of the Hawaii International Conference on System Sciences, 2000.
[19] S. Lindsey, C. Raghavendra, K. Sivalingam, Data gathering algorithms in sensor networks using energy metrics, IEEE Transactions on Parallel and Distributed Systems 13 (9) (2002).
[20] Y. Ossama, S. Gahmy, Heed: a hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks, IEEE Transactions on Mobile Computing 3 (4) (2005).
[21] ns2 Network Simulator, <http://www.isi.edu/nsnam/ns/>.
[22] XBOW, 2nd generation micamote, <http://www.xbow.com>.

Albert Harris currently holds a post doctorate position at the University of Illinois at Urbana-Champaign. He completed his PhD from UIUC in 2006 under Robin Kravets with a main focus on energy consumption in wireless networks. He completed a one year post doctorate position at the University of Padova, Italy under Michele Zorzi. His current interests are underwater acoustic networks, delay tolerant networks, and network support for assisted living and disaster recovery environments.

Robin Kravets is currently an associate professor at the Computer Science Department at the University of Illinois, Urbana-Champaign. Dr. Kravets received her Ph.D. from the College of Computing, Georgia Institute of Technology in 1999. She is the head of the Mobius group at UIUC, which researches communication issues in mobile, ad hoc and sensor networking, including power management, connectivity management, transport protocols, admission control, location management, routing and security. Her research has been funded by various sources, including the National Science Foundation, HP Labs and Boeing. For a list of publications and more detailed information, please visit: http://mobius.cs.uiuc.edu.

Indranil Gupta is Assistant Professor of Computer Science in the University of Illinois at Urbana-Champaign. He completed his PhD from Cornell University in 2004, and received the NSF CAREER award in 2005. Indranil heads the Distributed Protocols Research Group (DPRG) at UIUC, which studies various distributed systems such as peer-to-peer systems, the Grid, sensor networks, and design methodologies that span the breadth of these areas. He has served on the Program Committees for several ACM and IEEE conferences.