Energy-Efficient Data Aggregation Hierarchy for Wireless Sensor Networks

Yuanzhu Peter Chen   Arthur L. Liestman   Jiangchuan Liu*

School of Computing Science
Simon Fraser University
8888 University Drive, Burnaby, British Columbia, Canada
{yzchen, art, jcliu}@cs.sfu.ca

Abstract

A network of sensors can be used to obtain state-based data from the area in which they are deployed. To reduce costs, the data, sent via intermediate sensors to a sink, is often aggregated (or compressed). This compression is done by a subset of the sensors called aggregators. *Since sensors are usually equipped with small and unreplenishable energy reserves, a critical issue is to strategically deploy an appropriate number of aggregators so as to minimize the amount of energy consumed by transporting and aggregating the data.*

In this paper, we first study single-level aggregation and propose an Energy-Efficient Protocol for Aggregator Selection (EPAS). Then, we generalize it to an aggregation hierarchy and extend EPAS to a Hierarchical Energy-Efficient Protocol for Aggregator Selection (hEPAS). We derive the optimal number of aggregators with generalized compression and power-consumption models, and present fully distributed algorithms for aggregator deployment. Simulation results show that our algorithms significantly reduce the energy consumption for data collection in wireless sensor networks. Moreover, the algorithms do not rely on particular routing protocols, and are thus applicable to a broad spectrum of application environments.

1 Introduction

A wireless sensor network is a collection of sensors interconnected by wireless communication channels. Each sensor node [1] is a small device that can collect data from its surrounding area, carry out simple computations, and communicate with other sensors or with the controlling authorities of the network. Long distance communications are achieved in a multi-hop fashion. Such networks have been realized due to recent advances in micro-electro-mechanical systems and are expected to be widely used for such applications as environment monitoring, intrusion detection, and earthquake warning [28].

In many of these applications, the data to be collected is *state-based*, that is, it consists of measurements of ambient surroundings. Significant redundancies often exist in such data due to spatial-temporal correlations. These local redundancies can be removed prior to sending the over-sized raw data to the sink and draining the limited sensor energy store. This process, referred to as *data aggregation* or *data fusion*, is quite attractive as it is often infeasible or costly to replenish the batteries of the deployed sensors.

It is worth noting, however, that the amount by which the data size may be reduced by aggregation depends on the application. For example, simple statistical values such as sum, mean or deviation, can be easily aggregated into a single scalar or vector. On the other hand, a temperature map of a region would allow more limited reduction. For example, a wavelet scheme separately computes the wavelet transform for each subregion first, and then merges the resulting wavelet coefficients of subregions to obtain the wavelet transform of the entire region. Here, the number of wavelet coefficients of a subregion may increase as does the number of sensors therein.

In this paper, we investigate energy-efficient aggregator deployment in wireless sensor networks. A unique feature of our study is that we consider a general compression model for data aggregation, which is more realistic than the *infinite compression* [11, 3, 18] allowed in previous studies. We begin by using only a single level of aggregation. With this restriction, we calculate the number of aggregators needed to minimize the amount of total energy consumed in the network. A practical Energy-efficient Protocol for Aggregator Selection (EPAS) is presented to

achieve the target number of aggregators. Next, we demonstrate that multiple levels of aggregation can further reduce energy consumption. EPAS is then extended to *h*EPAS (Hierarchical EPAS) to provide a multiple level solution. We give fully distributed algorithms for aggregator deployment in the above protocols, which are applicable to a broad spectrum of state-based data collection applications in sensor networks.

The performance of these algorithms are examined by experiments. Our results demonstrate that EPAS conserves energy consumed both by the entire network and by the most heavily loaded sensors. The energy consumption can be further reduced by using *h*EPAS. The number of levels in the hierarchy is also a critical factor, and our results provide a general guideline toward desirable settings of the aggregation levels.

The remainder of this paper is organized as follows. In Section 2, we provide some background and review related work. An energy-efficient protocol for aggregator selection (EPAS) for one-level data aggregation is presented in Section 3. In Section 4, we generalize EPAS to accommodate aggregation hierarchies, called *h*EPAS. The performance of these protocols is examined in Section 5. Finally, Section 6 concludes the paper and offers some future research directions.

2 Background and Related Work

Wireless sensor networks have received much attention due to the number of potential applications of this technology. Many data communication protocols have been proposed lately, such as DD [12], TAG [16], TTDD [25], GRAB [26], Pilot [15], LEACH [11], and LAF [20]. Recent surveys by Akkaya and Younis [1], by Akyildiz *et al*. [2], and by Tilak, Abu-Ghazaleh and Heinzelman [23] include information on these and other protocols. There are three types of data collection in sensor networks. *Event-based data*, such as intrusion detection or object tracking, is collected when an event at a particular venue within the deployment region occurs. The event is confirmed by detecting sensors using local consensus and reported to the control authority [6, 19]. *Focused state-based data* is collected in response to a query sent to selected sensors requesting relevant data [12, 26]. *Global state-based data*, such as temperature or humidity, is collected by sensors all over the deployment area and transmitted to the sink [17]. Our interest here is in global state-based data.

Bhardwaj, Garnett, and Chandrakasan [4] provided an upper bound on the lifetime of sensor networks that are engaged in event detections. In their model, the energy consumed for a node to relay (that is, to receive and transmit) a unit of data to another node at distance d is denoted by $\alpha_1 + \alpha_2 d^l$, where α_1 and α_2 depend on the hardware im-

plementation of the sensors and l is the path attenuation exponent (usually in the range $2 \leq l \leq 4$). The distance from one sensor to the next that minimizes the energy consumed is the *characteristic distance*, denoted d_{char}, where $d_{char} = \sqrt[l]{\frac{\alpha_1}{\alpha_2(l-1)}}$. This value depends only on the hardware design specifications and the environment.

For state-based data collection, Heinzelman, Chandrakasan and Balakrishnan [11] presented a clustering algorithm (LEACH) to aggregate the data from sensors. In LEACH, each sensor becomes a *clusterhead* with a fixed probability during startup and every non-clusterhead sensor joins the cluster of a nearest clusterhead. The clusterheads act as aggregators. As clusterheads are likely to consume more energy than non-clusterheads, LEACH allows rotation of clusterhead status. Alternatively, unequal-sized clusters can be used to balance the sensor energy consumption [21].

Mhatre and Rosenberg [18] consider two types of nodes: regular sensors (type 0) and more powerful sensors (type 1) that can serve as clusterheads. Their work focuses on determining the numbers of type-1 sensors in a single aggregation level. A hierarchical clustering algorithm is proposed by Bandyopadhyay and Coyle [3]. In their model, they calculate the number of aggregators in each level for energy conservation. Clustering sensors and mobile ad hoc nodes, in general, has been an intensively studied area, see [5, 6, 7, 8, 13, 22, 24, 27].

Our hierarchical aggregator selection protocol (*h*EPAS) is motivated by these studies. In particular, Bandyopadhyay and Coyle [3] show that the use of a hierarchical structure can help conserve energy. However, we consider a more realistic compression model, and propose a general hierarchical framework to minimize the total energy consumed by both communication and aggregation.

3 One Level Aggregation

We begin by allowing only one level of data aggregation. Sensor nodes are partitioned into clusters, each with a clusterhead. The sensors within each cluster periodically send their data to the clusterhead. The clusterhead compresses the data collected from all members and sends the aggregated data to the sink. We first construct an ideal model, where the sensors and the aggregators are uniformly distributed over the region. Then we present a distributed algorithm, EPAS, to select the aggregators under practical constraints.

3.1 System Model and Notations

We first state our assumptions and introduce some notation to be used. A summary list of the notation used in this paper is given in Table 1.

Symbol	Meaning
a	Radius of the network deployment region
a_i	Radius of a level i cluster
b	Coverage radius of an aggregator
c	Data compression overhead
d_{char}	Characteristic distance, $\sqrt[l]{\frac{\alpha_1}{\alpha_2(l-1)}}$
E_a	Energy consumed by aggregating data in a single cycle in single-level context
E_{ai}	Energy consumed by aggregating data in a single cycle in level i
E_{ci}	Energy consumed by transporting data from level i to level $i+1$ in a single cycle
$f_a(\cdot)$	Aggregation energy consumption function
$g(\cdot)$	Data compressibility function
h	Number of levels in the hierarchy
k	Number of aggregators in single-level context
k_i	Number of aggregators in level i
l	Propagation loss exponent
n	Number of sensors in the network
r	Sensor data rate
r_i	Level i aggregator data rate
α_1	Circuit energy consumption coefficient
α_2	Antenna energy consumption coefficient
α	$\frac{\alpha_1+\alpha_2 d_{char}^l}{d_{char}}$
β	Aggregation energy consumption coefficient
γ	Data compression ratio

Table 1. List of notation.

Consider a network of n sensors uniformly distributed over a region. Many large-scale sensor networks such as environment monitoring sensors dropped from aircraft have this property [2, 28]. Route calculation is carried out during the initial setup after the aggregators (clusterheads) are selected, and the sensors send packets to their respective clusterheads using multi-hop paths (if necessary). Each hop in these paths is roughly of characteristic distance d_{char} [4]. That is, each node forwards the data to a node that is approximately d_{char} closer to the destination.

We assume that data collection is synchronized by cycles, where each cycle consists of a round of data collection, transmission, and aggregation. During each cycle, sensors collect data. The data generated is then sent to the aggregator as a packet of r bits. Each aggregator compresses the data it receives from the sensors of its cluster and then forwards the data to the sink. We assume that, by relaying packets via hops of the characteristic distance, transporting one unit of data a distance d consumes $\alpha \times d$ units of energy, where $\alpha = \frac{\alpha_1+\alpha_2 d_{char}^l}{d_{char}}$ [4]. That is, to send one unit of data a distance one requires the sender to expend $\alpha_2 d_{char}^l$ units of energy to transmit the message. The sender and receiver together use a total of α_1 units of energy internally.

We use a general function $g(x)$ to represent the data compressibility at aggregators. Basically, $g(x)$ gives the data volume after compression as a function of the input data volume x. When $g(x)$ is a constant, this is the infinite compression assumed in many of the previous studies [11, 3]. We use $f_a(x)$ to denote the energy consumed by compressing x data units. This is generally proportional to x but need not be. Although in practice the energy consumed by compressing a unit of data may be significantly less than that consumed by transmitting it (see [28]), we include this cost in our model for completeness, e.g., to accommodate advanced algorithms like Wavelet compression [28].

3.2 Optimal Number of Aggregators

We would like to determine the number of clusterheads (aggregators) that minimizes the total energy consumed by transmitting and aggregating data under our model. For simplicity, we assume that the sensors are deployed in a circular region A of radius a meters with the sink located at the center of the circle. Our solution can be easily extended to accommodate other region shapes or sink locations.

Let E_{c0} denote the total energy consumed by all of the sensors sending data to their respective aggregators in a single cycle. Consider the area covered by cluster C centered at (x_c, y_c). The total distance that the data packets travel from all members of C to (x_c, y_c) is $\frac{n}{\pi a^2} \times \iint_{(x,y)\in C} \sqrt{(x-x_c)^2 + (y-y_c)^2}\, dxdy$, where $\frac{n}{\pi a^2}$ is the sensor density.

If each sensor chooses the closest aggregator as its clusterhead, the sensors essentially form a Voronoi diagram of the network region where each cluster corresponds to a Voronoi cell. For large k, a typical cluster can be approximated as a circle of radius $\frac{a}{\sqrt{k}}$ [14] [2] with the aggregator at the center. With this, the above expression evaluates to $\frac{2an}{3k^{\frac{3}{2}}}$. Thus, after factoring in the α coefficient to obtain the energy consumption, the sensor data rate r, and summing over the k aggregators, we have

$$E_{c0} = \frac{2\alpha anr}{3k^{\frac{1}{2}}}. \tag{1}$$

Let E_a denote the total energy consumed by data aggregation in a single cycle. Since the aggregator receives data at an average rate of $\frac{nr}{k}$ bits per cycle, we have

$$E_a = k \times f_a\left(\frac{nr}{k}\right). \tag{2}$$

Let E_{c1} denote the total energy consumed by all of the aggregators sending this data to the sink in a single cycle.

---

[2]More details on the problem of covering geometric spaces with spheres can be found in Conway and Sloane [10].

Since the data is sent by an aggregator at a rate of $g\left(\frac{nr}{k}\right)$ bits per cycle and the aggregator density is $\frac{k}{\pi a^2}$, we have $E_{c1} = g\left(\frac{nr}{k}\right) \times \alpha \times \frac{k}{\pi a^2} \times \iint_{(x,y)\in A} \sqrt{x^2 + y^2}\, dx dy$, which evaluates to

$$E_{c1} = g\left(\frac{nr}{k}\right) \times \frac{2k\alpha a}{3}. \qquad (3)$$

Summing up Equations (1), (2) and (3), we have

$$\frac{2\alpha a n r}{3k^{\frac{1}{2}}} + k \times f_a\left(\frac{nr}{k}\right) + g\left(\frac{nr}{k}\right) \times \frac{2k\alpha a}{3}. \qquad (4)$$

Consider a typical circuit power-consumption model where the aggregation energy consumption is proportional to the volume of the data to be compressed, that is, $f_a(x) = \beta x$, for some constant β. Also consider a typical linear compression model, $g(x) = \gamma x + c$, where γ $(0 \leq \gamma \leq 1)$ is the compression ratio and c is the compression overhead [9]. It follows that the number of aggregators that minimizes the energy consumption of the network in a single cycle (that is, Equation (4)), is $k = \left(\frac{nr}{2c}\right)^{\frac{2}{3}}$. Substituting $g(x) = \gamma x + c$ into Equation (3), we see that the energy consumed by transporting the data from k aggregators to the sink is proportional to $nr\gamma + ck$. Thus, the contribution of γ to the value of Equation (3) (and (4)) is independent of k.

Thus, to minimize energy consumption, there should be $\left(\frac{nr}{2c}\right)^{\frac{2}{3}}$ aggregators. This conclusion also applies to the special case of $\beta = 0$, that is, where the energy required to compress data is negligible (as assumed in many existing sensor systems [28]).

3.3 Distributed Aggregator Selection – EPAS

In this section, we propose a practical Energy-Efficient Protocol for Aggregator Selection (EPAS) that follows our optimal solution in the previous subsection.

EPAS is a randomized and fully distributed algorithm that consists of two phases. In the first phase, each sensor chooses to be a clusterhead (aggregator) with probability p_1 independently for some $p_1 \in [0, \frac{k}{n}]$. Suppose that each clusterhead has a fixed coverage radius of b meters. (See Section 3.4 for discussion of determining the value of b.) In the second phase, each sensor that is not within the coverage radius of some clusterhead declares itself to be a clusterhead with probability p_2.

By careful choice of p_1 and p_2, we can ensure that the expected number of aggregators is k. To do that, we leverage the following propositions.

Lemma 1 *After phase 1 of EPAS, the probability that a sensor c is not covered is* $(1 - p_1)\left(1 - \frac{p_1 b^2}{a^2}\right)^{n-1}$, *where p_1 is the phase-1 selection probability, and b and a are the coverage and network region radii, respectively.*

Proof. Let X be a random variable denoting the number of sensors other than c that are contained in a circle of radius b centered at c. Thus, we have $Prob[X = i] = \binom{n}{i} \left(\frac{b^2}{a^2}\right)^i \left(1 - \frac{b^2}{a^2}\right)^{n-i-1}$, for $i = 0, 1, \ldots, n-1$. Further, $(1 - p_1)^X$ is a random variable indicating the probability that c is not covered by any selected clusterhead after phase 1. Its expected value is

$$
\begin{aligned}
E\left[(1 - p_1)^X\right] &= \sum_{i=0}^{n}(1 - p_1)^i \times Prob[X = i] \\
&= \sum_{i=0}^{n}(1 - p_1)^i \binom{n}{i} \left(\frac{b^2}{a^2}\right)^i \left(1 - \frac{b^2}{a^2}\right)^{n-i-1} \\
&= \left(1 - \frac{p_1 b^2}{a^2}\right)^{n-1}.
\end{aligned}
$$

This is also the probability that c is not covered by any selected clusterhead. Factoring in the probability that c does not select itself, we have the expected probability that sensor c is not covered: $(1 - p_1)\left(1 - \frac{p_1 b^2}{a^2}\right)^{n-1}$. $\square$

For the expected number of clusterheads generated by EPAS to be k, the selection probabilities of the two phases should satisfy the following condition.

Theorem 2 *The expected number of clusterheads generated by EPAS is k iff p_1 and p_2 are chosen such that* $p_1 + p_2(1 - p_1)\left(1 - \frac{p_1 b^2}{a^2}\right)^{n-1} = \frac{k}{n}$.

Proof. Let random variable Y denote the number of clusterheads generated by EPAS. Let binary random variable Y_i be 1 if and only if sensor c_i $(i = 1, 2, \ldots, n)$ becomes a clusterhead. Thus we have $E[Y_i] = \frac{E[Y]}{n} = \frac{k}{n}$. Let binary random variables $Y_{i,1}$ and $Y_{i,2}$ denote the fact that c_i becomes a clusterhead in phases 1 and 2, respectively. Since the events that c_i becomes a clusterhead in either phase are mutually exclusive, we have

$$
\begin{aligned}
E[Y_i] &= Prob[Y_i = 1] \\
&= Prob[(Y_{i,1} = 1) \cup (Y_{i,2} = 1)] \\
&= Prob[Y_{i,1} = 1] + Prob[Y_{i,2} = 1] \\
&= E[Y_{i,1}] + E[Y_{i,2}].
\end{aligned}
$$

We know that $E[Y_{i,1}] = p_1$. In addition, we have $E[Y_{i,2}] = p_2 \times Prob[c \text{ is not covered in phase 1}]$. Based on Lemma 1, we have $Prob[c \text{ is not covered in phase 1}] = (1 - p_1)\left(1 - \frac{p_1 b^2}{a^2}\right)^{n-1}$. As a result, we have $\frac{k}{n} = E[Y_{i,1}] + E[Y_{i,2}] = p_1 + p_2(1 - p_1)\left(1 - \frac{p_1 b^2}{a^2}\right)^{n-1}$. $\square$

After phase 2, the expected number of aggregators is k. Each sensor that is not an aggregator selects the closest aggregator as its clusterhead. Thus, the clusters essentially form a Voronoi partitioning of the network.

IEEE
COMPUTER
SOCIETY

3.4 Discussion

Given the number of sensors, deployment area, compression ratio, characteristic distance, and other network parameters, we can calculate the optimal number of aggregators k. Given this target number of aggregators, we can choose appropriate probabilities p_1 and p_2 offline. After deployment, the sensors select k aggregators. Each of the aggregators broadcasts its status to all sensors within coverage radius b. For large k, k circles of radius $b = \sqrt{\frac{3\sqrt{3}}{2\pi}}\frac{a}{\sqrt{k}} \approx 1.0996 \times \frac{a}{\sqrt{k}}$ can cover the entire region of area πa^2 [14]. We use a larger coverage radius $b = \frac{2a}{\sqrt{k}}$ to ensure that most of the sensors are within the coverage radius of at least one aggregator while keeping the broadcast radius small. In Section 5, we report on a set of experiments to determine how many sensors lie outside the coverage areas for k aggregators with coverage radius $b = \frac{2a}{\sqrt{k}}$.

4 Hierarchical Aggregation

We now consider a more general framework that organizes the aggregators in a hierarchy. We begin with all sensors in level 0 of the hierarchy. From those sensors, we select a subset as aggregators for level 1. From the level 1 aggregators, we select a subset to act as level 2 aggregators. Similarly, we select a subset of the aggregators at each level to act as aggregators at the next higher level. Finally, the sink (which may not be an aggregator of any of the other levels) is the only aggregator of level $h + 1$.

Once the aggregation hierarchy is established, sensors collect data and send it to the nearest level 1 aggregator. The level 1 aggregators collect this data from their sensors, aggregate it, and forward it to the nearest level 2 aggregator. This process continues until the level h aggregators forward the data to the sink.

In this section, we modify the method of Section 3 to determine the optimal number of aggregators in each level of the hierarchy. Then we extend EPAS to hEPAS, its hierarchical version.

4.1 Optimal numbers of aggregators in the hierarchy

We denote the number of aggregators in level i by k_i ($i = 0, 1, ..., h + 1$). Note that $k_0 = n$ and $k_{h+1} = 1$. The data is sent out of a level i aggregator to its clusterhead at a rate of r_i bits/cycle and $r_0 = r$. As before, the data compression rate is described by function $g(\cdot)$. We assume that aggregators in each level are distributed uniformly. That is, a level i aggregator receives data from $\frac{k_{i-1}}{k_i}$ level $(i - 1)$

aggregators. The data rate r_i can be expressed as:

$$r_i = \begin{cases} r & \text{if } i = 0 \\ g\left(r_{i-1} \times \frac{k_{i-1}}{k_i}\right) & i = 1, 2, \ldots, h. \end{cases}$$

Let E_{ai} be the total energy consumed by the compression done by all of the aggregators of each level i in a single cycle. We have $E_{ai} = k_i \times f_a\left(\frac{k_{i-1}}{k_i} \times r_{i-1}\right)$.

We now consider E_{ci}, the total energy consumed by transporting data from level i aggregators to level $(i + 1)$ aggregators ($i = 0, 1, \ldots, h$) in a single cycle. As before, a typical level $(i + 1)$ cluster C can be approximated with a circle of radius $a_{i+1} = a/\sqrt{k_{i+1}}$, centered at (x_c, y_c), and the density of the level i aggregators is $\frac{k_i}{\pi a^2}$. The portion of E_{ci} within this cluster is $r_i \times \alpha \times \frac{k_i}{\pi a^2} \times \iint_{(x,y) \in C} \sqrt{(x - x_c)^2 + (y - y_c)^2}\, dx dy = \frac{2\alpha a k_i r_i}{3k_{i+1}^{\frac{3}{2}}}$. Therefore, summing over the k_{i+1} level $(i + 1)$ clusters, we have $E_{ci} = \frac{2\alpha a k_i r_i}{3k_{i+1}^{\frac{1}{2}}}$.

Thus, the total energy consumed in a single cycle is

$$\sum_{i=1}^{h} E_{ai} + \sum_{i=0}^{h} E_{ci}, \tag{5}$$

which is a function of the k_i.

Given values of r, a, α, γ and c for a particular system, the values of k_i minimizing the above total energy consumption can be calculated. These values can then be used to configure the aggregator selection protocol.

4.2 Hierarchical EPAS

Here, we propose a Hierarchical Energy-Efficient Protocol for Aggregator Selection (*h*EPAS). We assume that the optimal number of aggregators in each level i is k_i ($i = 1, 2, \ldots, h$) as calculated above. The protocol selects an expected k_i sensors as the level i aggregators. *h*EPAS executes for h iterations. Each iteration is similar to EPAS (Section 3.3). During iteration i, a level $(i - 1)$ aggregator chooses to become a level i aggregator with probability $p_i \in [0, \frac{k_i}{n}]$ in the first phase. Each chosen aggregator has a coverage radius of $b = \frac{2a}{\sqrt{k_i}}$. In the second phase, a level $(i - 1)$ aggregator that is not covered by any level i aggregator chooses to become a level i aggregator with probability p_2, where p_1 and p_2 satisfy the condition described in Theorem 2 for $k = k_i$. After the aggregators of all levels are chosen, each level i aggregator ($i = 0, 1, \ldots, h$) joins the cluster of the nearest level $(i + 1)$ aggregator.

5 Performance Evaluation

We evaluate the performance of EPAS and *h*EPAS through simulations using our custom simulator. We first

investigate the energy consumption versus the number of aggregators in the single-level case in order to understand the effectiveness of EPAS. Then we study the energy saving gained by adding more levels to the aggregation hierarchy.

Parameter	Value
Number of sensors n	10000
Network radius a	1000 m
Communication circuit power α_1	5×10^{-8} J/bit
Communication antenna power α_2	1×10^{-10} J/bit/m^2
Propagation loss exponent l	2
Characteristic distance d_{char}	22.36 m
Aggregation energy consumption rate β	5×10^{-8} J/bit
Sensor data rate r	160 bits/sec
Data compression ratio γ	25%
Data compression overhead c	32 bits

Table 2. System specifications.

In the simulation, we use a network of 10000 sensors uniformly deployed in a circular region of radius 1000 meters. The system specifications we use are similar to those used by Heinzelman, Chandrakasan and Balakrishnan [11] and by Mhatre and Rosenberg [18] (see Table 2). We assume that the sensors sample the environment every minute. This differs slightly from the assumptions in these previous paper but is reasonable for our application. The measured values are converted to 16-bit digital representations and a single cycle lasts for 10 minutes. Therefore, the sensor data rate is 160 bits/cycle.

The regions served by the aggregators of each level form a Voronoi diagram of the overall network region. In the single level EPAS protocol, each aggregator compresses the data collected from sensors located within its Voronoi cell and sends it to the sink at the center of the region. To mitigate the impact of inefficient routing, we assume an idealized routing protocol, Characteristic Distance Progressive Routing (CDPR), to approximate straight-line routing [4]. In CDPR, when a packet intended for node d is at node s with $dist(s, d) > d_{char}$, it is forwarded to an intermediate node v_1 that is the closest to the target point t, a point on the line segment (s, d) at distance d_{char} from s. The packet is forwarded one step at a time until it reaches node v_i within distance d_{char} of d from which it is forwarded to d directly.

Our next goal is to choose a suitable value of p_1, and thus that of the corresponding p_2, that leaves few sensors uncovered after EPAS. To do this, we fix k and n, select a coverage radius b, and then, using various appropriate values of p_1, simulate the placement of the sensors and the choice of aggregators using that value of p_1. For each simulation, we count the number of uncovered sensors. In particular, we then assume that there are 10000 sensors and we choose the expected numbers of aggregators k to be 25, 100, 400, or 1600. Having fixed k, we use the coverage radius $b =$
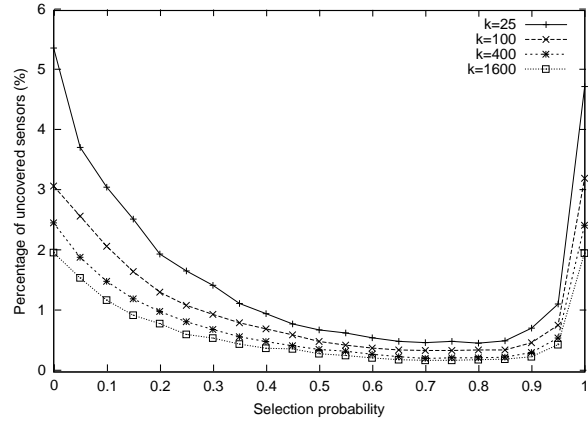

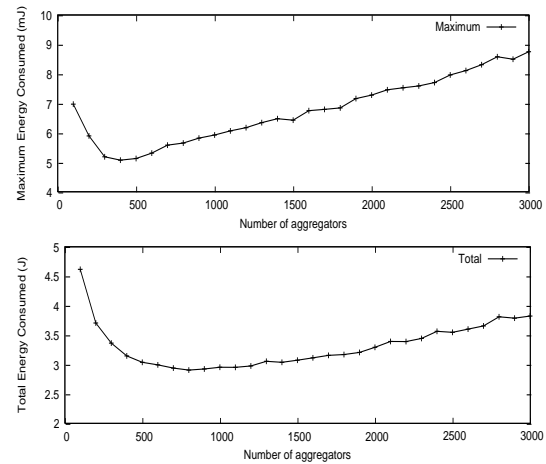**Figure 1. Uncovered sensors vs phase-1 probability.**

Figure 2. Single-level energy consumption.

$\frac{2 \times 1000}{\sqrt{k}}$. For the above values of k, this gives $b = 400, 200, 100$, and 50 meters, respectively. We consider each value of p_1 such that $p_1/(k/n) = 0.05i$, where $i = 0, 1, 2, \ldots, 20$. Figure 1 shows the percentage of sensors not covered by any aggregator for each choice of $p_1/(k/n)$. That is, the percentage of sensors not covered by any aggregator varies as p_1 does. Moreover, the fewer uncovered sensors, the better an aggregator's Voronoi cell can be approximated by a circle. Thus, choosing $p_1/(k/n) \in [0.7, 0.8]$ should yield the fewest uncovered sensors. For subsequent experiments, we use 0.75 as the value for $p_1/(k/n)$ and obtain the value of p_2 from Theorem 2.

We now measure the energy consumed in a single cycle of the data collection assuming a single level of aggregation. We fix the number of aggregators, select them, and then simulate the collection of data at each sensor, transmission
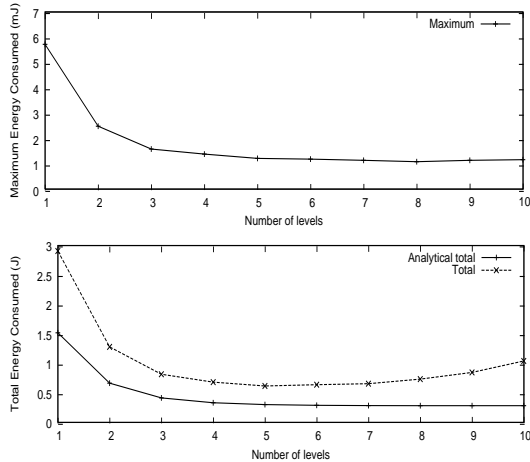
Figure 3. Hierarchical energy consumption.

of the data to the aggregators, aggregation of the data, and transmission of the aggregated data to the sink. For each simulation, we compute the maximum energy consumed by an individual sensor in the network. That is, if the sensor is not an aggregator, we count the energy needed to transmit its data to the aggregator and any energy used to forward other data during the cycle. For aggregators, we also include the cost of aggregation. In addition to determining the maximum of these costs, we also compute the total of these costs over all sensors. In the experiments, we again assume that we have 10000 sensors. We choose the number of aggregators to be a value $100j$, where $j = 1, 2, \ldots, 30$. Figure 2 shows the resulting maximum energy consumption and total energy consumption values. The total energy consumption appears to be minimized when the number of aggregators is between 800 and 900. Note that the value predicted by Equation (3) is 855. The maximum energy consumption is minimized when the number of aggregators is between 450 and 500 and is reasonably small for the range between 300 and 1000. Our experiments show that these most heavily loaded nodes are generally located near the aggregators or the sink. Such nodes are required to transport large amounts of data since the data is being concentrated at the aggregators and the sink.

We now consider whether additional energy savings can be achieved by instituting a hierarchical structure for the aggregators using the hEPAS protocol as described above. In an h-level hierarchy, we have calculated the number of aggregators k_i at each level i of the hierarchy that minimize the total energy consumption in an ideal situation. This total is given by Equation (5) and we can calculate the values of k_i that achieve this minimum. Table 3 shows the values of k_i for h levels where $1 \leq h \leq 10$ with a total of 10000 sensors deployed. We performed some experiments for h-level hierarchies where $1 \leq h \leq 10$ using the computed values k_i

shown in the table. In each experiment, we used the hEPAS protocol to select k_i aggregators at each level i and then measured the energy consumed by each sensor, recording both the maximum for any sensor and the total consumed by all sensors. For each choice of h, we performed 1000 experiments. The averages of these values over all of the experiments for each h are shown in the two plots in Figure 3. The ideal values from Equation (5) are included in the lower plot (which shows the total energy consumed). From these experimental results, we conclude that for 10000 sensors, the best choice would be to use a hierarchical solution with 5 levels. Note that the experimental results are reasonably close to the ideal values, which indicates that our ideal scenario is a good model of the actual situation.

6 Conclusion

In this paper, we studied the energy consumed in wireless sensor networks in which some sensors can aggregate the data. In particular, we considered one scenario that allowed a single level of aggregation assuming a general compressibility function. we calculated the number of aggregators needed to minimize the amount of total energy consumed in the network. A practical Energy-efficient Protocol for Aggregator Selection (EPAS) was presented to achieve the target number of aggregators.

We then considered a more general scenario with multiple levels of aggregation, and extended EPAS to its hierarchical version, hEPAS. We gave fully distributed algorithms for aggregator deployment in the above protocols, which are applicable to a broad spectrum of state-based data collection applications in sensor networks.

We performed a series of experiments, measuring energy consumption in networks with different numbers of levels and aggregators at each level. Overall, our experiments show that both the total energy consumption and the maximum energy consumption among sensors are significantly reduced by employing the proposed protocols.

This work can be extended in a number of different ways. For example, we can better balance the energy consumption among nodes by the use of mobile aggregators. We may also consider the problem of aggregator selection for sensor networks with heterogeneous nodes.

References

[1] K. Akkaya and M. Younis. A survey on routing protocols for wireless sensor networks. *Ad Hoc Networks*, 3 (3): 325–349, 2005.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38 (4): 393–422, 2002.

IEEE
COMPUTER
SOCIETY

h	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	E_t
1	855.0										1.5174
2	1848.8	369.8									0.6626
3	3210.9	808.4	177.7								0.4134
4	4478.2	1411.2	387.8	93.3							0.3306
5	5281.6	1953.8	659.4	198.3	52.6						0.2999
6	5689.9	2299.2	890.4	321.8	106.0	31.2					0.2878
7	5878.2	2478.7	1033.9	419.4	162.1	58.5	19.3				0.2829
8	5972.7	2565.3	1108.8	477.8	202.9	83.7	33.0	12.2			0.2810
9	5998.3	2600.7	1142.6	506.5	225.3	101.0	43.9	18.9	7.9		0.2804
10	6005.8	2608.4	1149.8	512.8	230.4	107.4	48.2	22.5	10.6	5.1	0.2803

Table 3. Number of aggregators.

[3] S. Bandyopadhyay and E. J. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proc. INFOCOM*, pp. 1713–1723, 2003.

[4] M. Bhardwaj, T. Garnett, and A. Chandrakasan. Upper bounds on the lifetime of sensor networks. In *Proc. IEEE Int. Conf. Comm.*, Vol. 3, pp. 785–790, 2001.

[5] J. Blum, M. Ding, A. Thaeler, and X. Cheng. Connected dominating set in sensor networs and MANETs. In D.-Z. Du and P. Pardalos, editors, *Handbook of Combinatorial Optimization*, Kluwer, 2004.

[6] W.-P. Chen, J. C. Hou, and L. Sha. Dynamic clustering for acoustic target tracking in wireless sensor networks. *IEEE Trans. Mobile Comp.*, 3 (3): 258–271, 2004.

[7] Y. P. Chen and A. L. Liestman. Approximating minimum size weakly-connected dominating sets for clustering mobile ad hoc networks. In *Third ACM Int. Symp. on Mobile Ad Hoc Networking and Comp. (MobiHoc'02)*, pp. 165–172, 2002.

[8] Y. P. Chen, A. L. Liestman, and J. Liu. Clustering algorithms for ad hoc wireless networks. In Y. Pan and Y. Xiao, editors, *Ad Hoc and Sensor Networks*. Nova, 2004.

[9] J. Chou, D. Petrovic, and K. Ramchandran. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks. In *Proc. INFOCOM*, pp. 1054–1062, 2003.

[10] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups (Grundlehren der mathematischen Wissenschaften)*, Springer, 1998.

[11] W. B. Heinzelman, A. Chandrakasan, and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wireless Comm.*, 1 (4): 660–670, 2002.

[12] C. Intanagonwiwat, R. Govindan, D. Estrin, J. S. Heidemann, and F. Silva. Directed diffusion for wireless sensor networking. *IEEE/ACM Trans. Networking*, 11: 2–16, 2003.

[13] L. Jia, R. Rajaraman, and T. Suel. An efficient distributed algorithm for constructing small dominating sets. In *Proc. 20th ACM Symp. Princ. Distr. Comp. (PODC'01)*, 2001.

[14] R. Kerschner. The number of circles covering a set. *American J. .Math.*, 61: 665–671, 1939.

[15] J. Luo, P. T. Eugster, and J.-P. Hubaux. Pilot: Probabilistic lightweight group communication system for ad hoc networks. *IEEE Trans. Mobile Comp.*, 3 (2): 164–179, 2004.

[16] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TAG: A tiny aggregation service for ad-hoc sensor networks. In *Proc. 5th Symp. OS Design and Implementation (OSDI)*, pp. 131–146, 2002.

[17] D. Marco and D. L. Neuhoff. Reliability vs. efficiency in distributed source coding for field-gathering. In *Proc. Inf. Proc. in Sensor Networks (IPSN)*, 2004.

[18] V. Mhatre and C. Rosenberg. Design guidelines for wireless sensor networks: Communication, clustering and aggregation. *Ad Hoc Networks*, 2 (1): 45–63, 2004.

[19] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Comm. ACM*, 43 (5): 51–58, 2000.

[20] H. Sabbineni and K. Chakrabarty. Location-aided flooding: An energy-efficient data dissemination protocol for wireless sensor networks. *IEEE Trans. Comp.*, 51 (1): 36–46, 2005.

[21] S. Soro and W. B. Heinzelman. Prolonging the lifetime of wireless sensor networks via unequal clustering. In *Proc. 5th Int. Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks (WMAN '05)*, 2005.

[22] J. Sucec and I. Marsic. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proc. INFOCOM*, pp. 1698–1706, 2003.

[23] S. Tilak, N. B. Abu-Ghazaleh, and W. B. Heinzelman. A taxonomy of wireless micro-sensor network models. *ACM SIGMOBILE Mobile Comp. and Comm. Review*, 6 (2): 28–36, 2002.

[24] J. Wu and W. Lou. Extended multipoint relays to determine connected dominating sets in MANETs. In *Proc. First IEEE Comm. Society Conf. on Sensor and Ad Hoc Comm. and Networks (SECON)*, 2004.

[25] F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang. A two-tier data dissemination model for large-scale wireless sensor networks. In *Proc. MobiCom*, pp. 148–159, 2002.

[26] F. Ye, G. Zhong, S. Lu, and L. Zhang. Robust data delivery protocol for large scale sensor networks. In *Proc. IEEE Int. Workshop Info. Proc. in Sensor Networks (IPSN)*, pp. 658–673, 2003.

[27] O. Younis and S. Fahmy. HEED: A hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks. *IEEE Trans. Mobile Comp.*, 3 (4): 366–379, 2004.

[28] F. Zhao and L. Guibas. *Wireless Sensor Networks : An Information Processing Approach*. Morgan Kaufmann, 2004.

IEEE COMPUTER SOCIETY