



Design guidelines for wireless sensor networks: communication, clustering and aggregation

Vivek Mhatre, Catherine Rosenberg *

School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285, USA

Received 15 June 2003; accepted 15 July 2003

Abstract

When sensor nodes are organized in clusters, they could use either single hop or multi-hop mode of communication to send their data to their respective cluster heads. We present a systematic cost-based analysis of both the modes, and provide results that could serve as guidelines to decide which mode should be used for given settings. We determine closed form expressions for the required number of cluster heads and the required battery energy of nodes for both the modes. We also propose a hybrid communication mode which is a combination of single hop and multi-hop modes, and which is more cost-effective than either of the two modes. Our problem formulation also allows for the application to be taken into account in the overall design problem through a data aggregation model.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Wireless sensor networks; Clustering; Single hop vs multi-hop; Data aggregation

1. Introduction

Wireless sensor networks are networks of wireless nodes that are deployed over an area for the purpose of monitoring certain phenomena of interest. The nodes perform certain measurements, process the measured data and transmit the processed data to a base station over a wireless channel. The base station collects data from all the nodes, and analyzes this data to draw conclusions about the activity in the area of interest. These networks are different from the traditional wireless

ad hoc networks, because the nodes in an ad hoc network are in general less energy constrained [1]. In ad hoc networks the communication paradigm is any-to-any, since any node may wish to communicate with any other node. However in most sensor networks the many-to-one communication paradigm is more common. This is because in case of sensor networks, nodes send their data to common sinks or cluster head nodes for processing. This many-to-one paradigm often results in non-uniform energy drainage patterns in the network.

In the context of ad hoc networks it is well-known that when the propagation loss exponent is high, multi-hop communication should be used to counter the high path loss. However when nodes are organized in clusters, and when they use multi-hop communication to reach the cluster head, the

* Corresponding author. Tel.: +1-765-494-0034; fax: +1-765-494-0880.

E-mail addresses: mhatre@ecn.purdue.edu (V. Mhatre), cath@ecn.purdue.edu (C. Rosenberg).

nodes closer to a cluster head have a higher load of relaying packets as compared to other nodes. When the nodes are mobile (as is the case in ad hoc networks), due to the randomness induced by the time varying node positions, this relaying load gets (more or less) evenly distributed over all the nodes. However in most sensor networks nodes are static. Consequently the nodes closer to the cluster head get overburdened constantly. On the other hand when the nodes use single hop communication to reach the cluster heads, the nodes located farther away from a cluster head have the highest energy burden due to long range communication. The cluster heads themselves have the extra burden of performing long range transmissions to the distant base station.

The problem we address is that of determining the optimum number of cluster heads, of dimensioning, and determining the battery energy of the nodes, and determining the optimum mode of communication in each cluster (single hop or multi-hop). Most of the work in the sensor network literature assumes one of the two modes (single hop or multi-hop) and then optimizes the system for that particular mode. However in our work we present a systematic cost-based comparison of the two modes for 1-D (linear), 2-D (planar) and 3-D (spatial) clusters. We also propose a model for data aggregation which serves as an entry point for the application in the overall network design problem that we study next. We then propose and analyze a hybrid communication mode which alternates between single hop and multi-hop modes to ensure a more uniform energy drainage pattern. In this study we mainly focus on the trade-offs involved between communication, clustering and aggregation, and hence it is difficult to account for all other aspects of sensor networks such as MAC and routing. Our objective is to study these trade-offs and provide guidelines to the sensor network designers. Our analysis pertains only to the data gathering sensor networks, and not to the event detection sensor networks. In data gathering networks the nodes periodically send their sensed data to the base station, while in event detection sensor networks the nodes are idle for long periods of time, and spring into activity only when the event of interest occurs.

The rest of the paper is organized as follows. In Section 2 we discuss some of the related work. Section 3 contains a brief outline of the problem statement and our approach. In Section 4 we study single hop versus multi-hop modes in a single cluster. In Section 5 we propose a new model for data aggregation and solve the overall system design problem with this model. In Section 6 we propose and study a hybrid mode of communication. Section 7 presents some case studies for some typical sensor network settings. Finally we conclude the paper in Section 8.

2. Related work

Bandyopadhyay et al. in [4] have studied a multi-hop clustered wireless sensor network. Nodes communicate with their respective cluster heads by using multi-hop communication. The cluster heads collect data from the nodes in their respective clusters, aggregate the gathered data, and send it to a base station located at the center of the region using multi-hop communication. For this scenario, the authors have provided expressions for the required cluster head densities in order to minimize the total energy expenditure in the network. However they do not provide any justification for choosing multi-hop mode for communication between the sensor nodes and the cluster heads, and between the cluster heads and the base station. Another point that needs to be stressed about the study in [4] is that the sensor nodes which are within one hop from the base station have excessive burden of relaying, and therefore when these nodes expire, connectivity is lost and the network becomes unusable. Clearly, instead of minimizing the total energy expenditure in the network, the goal should be to minimize the energy expenditure of the sensor nodes around the base station, since these nodes determine the lifetime of the system.

In [2], Heinzelman et al. study a clustered sensor network protocol called LEACH. The authors consider a scenario in which homogeneous, i.e., only one type of nodes are used, and the nodes communicate with their elected cluster heads using single hop communication. The cluster heads aggregate the received data, and transmit it to a dis-

tant base station using a single hop transmission. The authors present a distributed algorithm for cluster head selection. LEACH also uses rotation of the cluster heads for load balancing, since the cluster heads have the extra burden of performing the long range transmissions to the distant base station. Thus LEACH counteracts the problem of non-uniform energy drainage by role rotation.

In [8], the authors have studied the problem of designing a surveillance sensor network. Instead of using homogeneous nodes and cluster head rotation as in LEACH, the authors use two types of nodes; type 0 nodes which act as pure sensor nodes, and type 1 nodes which act as cluster heads. Cluster head nodes have higher hardware and software complexity as well as a higher battery energy requirement. Sensor nodes use multi-hop communication to send their data to the cluster heads. The authors formulate an optimization problem to minimize the overall cost of the network and determine the optimum number of cluster heads and the battery energies of both types of nodes.

In all the above studies [4,2,8] the authors use the following model for data aggregation. Irrespective of the size of the cluster, i.e., irrespective of the number of nodes in a cluster, the cluster head is assumed to aggregate the gathered data into a single packet whose length is fixed, and does not depend on the number of input packets. This model of infinite compressibility may be applicable to certain applications, but is not general enough to represent most sensor networks. We propose a more general and improved model for data aggregation and work with that model.

In [3], Bhardwaj et al. study a multi-hop sensor network. They provide an upper bound on the lifetime of the network by minimizing the energy spent on sending a packet from a source node to a destination node by using optimum number of relay nodes. However this analysis is not applicable to the scenarios in which a receiver node is located at the center of the region, because it does not take into account the fact that the nodes closer to the receiver have more packets to relay as compared to other nodes. The authors focus on one source-destination pair at a time without taking into account the many-to-one communication paradigm.

In [9], the authors provide bounds on the lifetime of a sensor network over all the collaborative data gathering strategies. For a given topology, there are several different routes that packets originating at a particular node can take to reach the destination node. These routes also include those paths in which the node does not necessarily communicate directly with its one hop neighbor. Instead, the node may transmit the packet directly to another node which is two or more hops away by spending more energy. Thus the total number of paths that a packet can take from source to destination grows exponentially as the number of nodes in the network increases. The authors determine the optimum fractions of time over which each of these paths should be sustained so as to minimize the network-wide energy expenditure. In order to obtain a polynomial time solution, the authors formulate the problem as a network flow problem which keeps things tractable. We discuss this work in more details in Section 6 where we discuss our hybrid communication mode.

3. Problem outline and our approach

We consider a region to be covered by sensor nodes. The number of sensor nodes is determined by the application requirements. Usually, each sensor node has a sensing radius and it is required that the sensor nodes provide coverage of the region with a high probability [8]. The sensing radius of each node depends on the phenomenon that is being sensed as well as the sensing hardware of the node. Thus in general the required number of sensor nodes is dictated by the application and hence we assume it to be a constant. We assume that the sensor nodes are randomly and uniformly distributed over the region. We also assume that the nodes are organized in clusters to take advantage of possible data aggregation at the cluster head nodes. The network is heterogeneous and there are two types of nodes; cluster head nodes and sensor nodes. The cluster head nodes act as the fusion points within the network. During each data gathering cycle, the sensor nodes send their sensed data to the closest cluster head node which performs data aggregation. Then the cluster head

directly transmits the aggregated data to a base station (assumed to be remotely located). The sensor nodes have simple functionality, since they perform sensing and relatively short range communication. However the cluster head nodes are more complex, since they co-ordinate MAC and routing within their cluster, perform data fusion, and perform long range transmissions to the remote base station.

The overall system design problem involves determining the optimum number of cluster head nodes, the optimum mode of communication within a cluster (single hop or multi-hop) and the required battery energies of both types of nodes. We formulate an optimization problem in which we associate a cost function with each type of node. The cost function takes into account the hardware cost and the battery cost of the node. We also take into account the data aggregation model, and then obtain an expression for the cost of the entire system for the two different communication modes within a cluster. Then we compare the minimized cost functions of both the solutions to determine the best solution. We break down this problem into smaller parts by first studying a typical cluster, since a cluster acts as a building block for the entire network.

4. Cluster design

In this section we study how the choice of the mode of communication within a cluster affects the required battery energy of the sensor nodes. The energy requirements of the cluster head nodes (aggregation energy and energy spent on communication with the base station) are later taken into account in the overall system design. Our analysis in the following subsections is restricted only to a single cluster, however we use these results later on in our analysis of the overall system.

4.1. System description

Consider a 2-D (planar) cluster. For simplicity, we assume that the cluster is a circular region and the cluster head is located at the center of this region. Let the radius of the cluster be a . There are N

sensor nodes uniformly distributed over the cluster area. During each data gathering cycle each sensor node senses and sends its sensed data to the cluster head. Aggregation is performed only at the cluster head node. All the sensor nodes are identical and have the same amount of initial battery energy. We would like to ensure that at least T data gathering cycles are possible until any of the nodes exhausts its battery. Or equivalently, we want to guarantee a lifetime of at least T units. We also assume that the cluster head is not energy constrained. (We consider the cluster head energy requirements in later sections.) We assume a simple communication model for the transceiver similar to the one in [2] in which the amount of energy required to transmit a packet over distance x is given by $l + \mu x^k$, where l is the amount of energy spent in the transmitter electronics circuitry, while μx^k is the amount of energy spent in the RF amplifiers to counter the propagation loss. Here μ takes into account the constant factor in the propagation loss term, as well as the antenna gains of the transmitter and the receiver. The value of the propagation loss exponent k is highly dependent on the surrounding environment. Usually, on-site measurements are performed to determine the value of k for a given site [7]. In environments such as buildings, factories and regions with dense vegetation, the value of k is high (3–5), while for free space the value of k is 2. When receiving a packet, only the receiver circuitry is invoked, and so the energy spent on receiving a packet is l . Thus to relay a packet over distance x , $2l + \mu x^k$ amount of energy is spent.

We assume that the cluster head nodes co-ordinate the MAC and the routing of packets within their clusters so that packet transmissions and receptions within each cluster are synchronized. Therefore there is no IDLE mode energy expenditure to account for. If the MAC protocol has any additional energy overheads, it is possible to take them into account by slightly modifying our analysis. However the approach to solving the problem remains the same. For simplicity we also assume that all the packets are of fixed length. Note that both l as well as μ are for a single packet, and hence the length of the packet has been absorbed in l and μ .

Usually the energy spent on sensing is small as compared to the communication energy. Besides it is equal to a constant multiplied by T for T data gathering cycles. After determining the energy spent on communication, we can simply add this constant amount of energy spent on sensing to obtain an exact expression for the total battery energy. This constant amount does not affect the choice of the communication mode (single hop or multi-hop) and therefore for simplicity we do not include it in our subsequent analysis.

4.2. Single hop mode

When the sensor nodes use single hop communication, there is no relaying of packets. Each node directly transmits its packet to the cluster head (see Fig. 1(a)). Since the communication is directly between the sensor nodes and the cluster head, only one node should transmit at a time, and a contention-less MAC is preferred and assumed. The lifetime¹ of the network is determined by the lifetime of the shortest-living node. In the case of a single hop network the sensor nodes located farthest from the cluster head (at a distance a) have to spend the maximum amount of energy. Since all the sensor nodes are alike, the dimensioning of the battery energy has to be performed with the worst case scenario in perspective. Hence in order to ensure a lifetime of at least T cycles, we require that the battery energy of the sensor nodes in the single hop communication system E_s be

$$E_s = T(l + \mu a^k). \quad (1)$$

Nodes may use power control to save energy and to reduce interference with the neighboring clusters. However this has no impact on the problem of battery dimensioning which needs to account for the worst case energy expenditure.

¹ We restrict ourselves to the definition of lifetime in which the first node expiration is taken to be the expiration of the sensor system. This is a conservative approach, especially for a system with single hop communication, since a single hop network continues to provide data updates even after the farthest nodes expire (although there are fewer updates), but a multi-hop network loses connectivity, and becomes non-functional after the nodes around the cluster head expire.

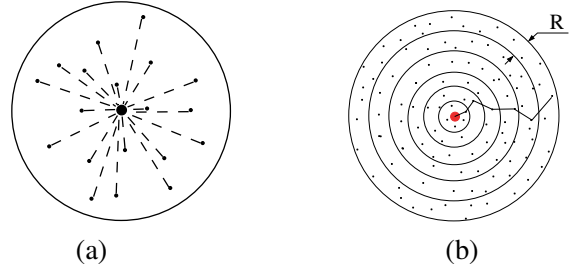


Fig. 1. Communication modes. (a) Single hop, (b) multi-hop.

4.3. Multi-hop mode

Now consider a cluster in which the sensor nodes reach the cluster head by using multi-hop communication. We assume a simple communication model in which each sensor node has a communication radius R over which it can communicate to reach its neighboring node. We also assume that $R < a$, i.e., the communication area of each node is smaller than the total area of the cluster. Otherwise the cluster is the same as the single hop communication cluster. For multi-hop communication to be possible it is necessary that R be sufficiently large so that the connectivity of the nodes is maintained. In [6] the authors have obtained a lower bound on the communication radius R in order to ensure connectivity of the nodes with a high probability. When n nodes are uniformly and randomly distributed over a unit area, the probability of connectivity of nodes is lower bounded by (Lemma 3.1, (1.15) of [6]),

$$P(\text{connectivity}) \geq 1 - ne^{-n\pi r^2(n)}.$$

This is a sufficient condition for connectivity, and therefore is a loose bound. To have node connectivity with a probability of at least $1 - \epsilon$, we have the following:

$$1 - ne^{-n\pi r^2(n)} \geq 1 - \epsilon \Rightarrow P(\text{connectivity}) \geq 1 - \epsilon$$

$$\Rightarrow r(n) \geq \sqrt{\frac{1}{n\pi} \log\left(\frac{n}{\epsilon}\right)}.$$

When N sensor nodes are distributed over an area πa^2 , using the above relationship, and scaling all the distances by the normalizing factor we obtain that R should be greater than or equal to r as in

(2), so that the nodes be connected with a probability of at least $1 - \epsilon$.

$$R \geq r = a \sqrt{\frac{1}{N} \log \left(\frac{N}{\epsilon} \right)}. \quad (2)$$

In this section we assume that $R \geq r$ so that multi-hop communication is indeed possible. We ignore the amount of energy spent on the routing updates and/or MAC control packets by assuming that the data traffic is much higher than the control traffic. For simplicity we also ignore the energy wasted during packet collisions as well as start-up transients.

In order to determine the worst case energy drainage in the network, we divide the circle into concentric rings of thickness R (see Fig. 1(b)). We note that with a multi-hop communication radius of R , if a packet is generated in the n th ring, during its journey to the cluster head, the packet has to travel through each of the inner rings. For each data gathering cycle, we determine the average energy expenditure of a sensor node in the n th ring, where n varies from 1 to a/R . Since the nodes are uniformly distributed, the average number of sensor nodes which lie outside the n th ring is $N(\pi a^2 - \pi(nR)^2)/\pi a^2$. Hence $N(\pi a^2 - \pi(nR)^2)/\pi a^2$ number of packets have to be relayed by the nodes in the n th ring into the $(n-1)$ th ring. There are $N(\pi(nR)^2 - \pi((n-1)R)^2)/\pi a^2$ nodes in the n th ring that have to relay the packets coming from the nodes outside the ring. If we denote the average number of packets that a typical node in the n th ring has to relay by k_n , then we obtain

$$\begin{aligned} k_n &= \frac{N(\pi a^2 - \pi(nR)^2)/\pi a^2}{N(\pi(nR)^2 - \pi((n-1)R)^2)/\pi a^2} \\ &= \frac{a^2 - n^2 R^2}{R^2(2n-1)}. \end{aligned} \quad (3)$$

In addition to relaying these k_n packets, the node also has to transmit its own packet. Hence the total average energy spent during one cycle by a node in the n th ring (denoted by e_n) is

$$e_n = (2l + \mu R^k)k_n + (l + \mu R^k).$$

To ensure a network lifetime of T , a node in the n th ring should have a battery energy of at least $E_m(nR)$ given by

$$E_m(nR) = T((2l + \mu R^k)k_n + (l + \mu R^k)). \quad (4)$$

For dimensioning the battery energy, we must consider the worst case energy drainage which corresponds to the maximum value of $E_m(nR)$ (i.e., k_n) over all the n values, and this corresponds to $n = 1$. This is something we would expect, since we know that the sensor nodes closest to the cluster head, i.e., the sensor nodes in the ring $n = 1$, have the highest relaying burden. Hence the required battery energy for the multi-hop scenario, E_m , is given by

$$\begin{aligned} E_m &= T((2l + \mu R^k)k_1 + (l + \mu R^k)) \\ &= T \left((2l + \mu R^k) \left(\frac{a^2}{R^2} - 1 \right) + (l + \mu R^k) \right). \end{aligned} \quad (5)$$

When $k = 2$, we obtain

$$E_s = T(l + \mu a^2), \quad (6)$$

$$E_m = T(l + \mu a^2) + 2Tl \left(\frac{a^2}{R^2} - 1 \right). \quad (7)$$

Since $R < a$, the second term in the expression for E_m is always positive. Thus we can see that $E_m > E_s$, i.e., the required battery energy is lower for single hop mode than multi-hop mode when $k = 2$. The reason being that the average number of packets to be relayed by a sensor node in the first ring, k_1 , scales as $(a^2 - R^2)/R^2 \approx 1/R^2$ while the energy required to relay each of these packets scales as μR^2 and hence the two terms balance each other in the product. In (7), the larger the R , the smaller the required energy, and this required energy is minimized when R is maximum which corresponds to the single hop scenario ($R = a$). On the other hand, when $k > 2$ the propagation loss term scales as μR^k , while the term corresponding to the average number of packets to be relayed still scales as $1/R^2$. As a result the choice of whether to use single hop or multi-hop mode when $k > 2$ depends on some other factors such as k , l , μ and the choice of R .

For a general $k > 2$, differentiating (5) and equating the result to 0 for minimizing E_m , the solution $R = \hat{R}$ is obtained as

$$\mu \hat{R}^k = \frac{4l}{k-2} \Rightarrow \hat{R} = \left(\frac{4l}{\mu(k-2)} \right)^{1/k}. \quad (8)$$

The \hat{R} thus obtained is independent of the dimensions of the region and depends only on the radio parameters k , l and μ . It can be shown that the second derivative of (5) is always positive. With \hat{R} as the radius of communication for multi-hopping the energy load on the nearby sensor nodes around the cluster head can be minimized.

However in general it may not be feasible to choose \hat{R} as the inter-hop distance. This is because the requirement of node connectivity for multi-hop communication imposes a lower bound on the communication radius of each node (2). If $r < \hat{R}$ then using \hat{R} as the inter-hop distance is feasible. If $r > \hat{R}$, then we do not have any choice but to use $R = r$, because with $R = \hat{R}$ node connectivity cannot be ensured and hence multi-hop communication cannot take place. Also note that if the size of the area is such that $a \leq \hat{R}$ then clearly single hop communication is the best solution. Hence the radius of communication that should be used for multi-hop communication, \tilde{R} , is given by

$$\tilde{R} = \min\{\max(r, \hat{R}), a\}. \quad (9)$$

Note that in (8), \hat{R} goes to 0 as l goes to 0 which suggests that because of the constant amount of energy that needs to be spent during relaying (2l), it is not always beneficial to use more and more intermediate hops. There is a trade-off involved and this trade-off was already pointed in [2,3]. However these studies do not take into account the fact that the energy load on the sensor nodes in a many-to-one communication paradigm varies depending on their location, and that it is the worst case load that determines the system lifetime. This is especially the case with multi-hop networks, because when the nodes closest to the cluster head expire, the network connectivity is lost. In single hop scenario the degradation is much less drastic, because nodes do not rely on each other to communicate with the cluster head.

The result in (8) is similar to the result obtained by Bhardwaj et al. in [3] where they define the *characteristic distance* as that distance which when used as the inter-node distance, minimizes the energy spent in sending a packet from a source node to a destination node. This characteristic distance, d_{char} , is

$$d_{\text{char}} = \left(\frac{\alpha_1}{\alpha_2(k-1)} \right)^{1/k}, \quad (10)$$

where α_1 is the constant energy spent during relaying which corresponds to $2l$ in our case, and α_2 corresponds to μ in our case. However note that in our case the denominator has a $k-2$ factor while in (10) this factor is $k-1$. The reason is that [3] do not take into account the fact that the relaying load scales as $1/R^2$. Although (8) and (10) look strikingly similar, note that they give considerably different results for typical values of k (between 2 and 5).

4.4. Multi-hopping in 3-D space and along a 1-D line

There are some applications in which the sensor nodes are deployed in 3-D space. For example sensor networks that monitor temperature in buildings, sensor networks for seismic measurements in structures, etc. In the previous sub-sections we confined ourselves to a 2-D scenario. But we can easily extend this analysis to 3-D space. We assume that nodes are uniformly distributed in the 3-D space. Just as we divided the circle of radius a into concentric rings of thickness R , we can divide the sphere of radius a into concentric shells of thickness R . The average relaying load on a node in the shell $n = 1$ is

$$k_1 = \frac{4\pi a^3/3 - 4\pi R^3/3}{4\pi R^3/3} \Rightarrow k_1 = \frac{a^3}{R^3} - 1.$$

Consequently (5) takes the following form:

$$E_m = T \left(\frac{2la^3}{R^3} - l + \mu R^{k-3} a^3 \right).$$

The corresponding \hat{R}_{3D} for the 3-D scenario is

$$\mu \hat{R}_{3D}^k = \frac{6l}{k-3} \Rightarrow \hat{R}_{3D} = \left(\frac{6l}{\mu(k-3)} \right)^{1/k}. \quad (11)$$

It can be shown that in the 3-D case single hopping is better than multi-hopping for $2 \leq k \leq 3$. The proof is exactly along the same lines as the 2-D case and therefore has been omitted.

We now consider the scenario in which sensor nodes are deployed along a line segment with

uniform distribution, and a cluster head is located at the midpoint of the segment. In this case the maximum relaying load varies as $1/R$. Using a similar approach as above we can prove that single hop communication is better than multi-hop communication when $k \leq 1$. Usually k is larger than one and therefore multi-hop communication is better than single hop communication. The optimum radius of communication \hat{R}_{ID} is then given by

$$\mu \hat{R}_{ID}^k = \frac{2l}{k-1} \Rightarrow \hat{R}_{ID} = \left(\frac{2l}{\mu(k-1)} \right)^{1/k}. \quad (12)$$

The above result is identical to (10), since we are considering relaying load in one dimension only.

5. Data aggregation and overall system design

So far we have studied the problem of dimensioning of the battery energy of sensor nodes by analyzing a single cluster. However when we study the problem of system design, we also need to address the problem of determining the optimum number of cluster heads. As it turns out this problem is related to the problem of battery energy dimensioning of the sensor nodes. This is because one of the parameters in (1) and (5) is a which is a measure of the size of each cluster. If the area of the region is fixed (say πA^2), then the size of each cluster is determined by the number of clusters. Thus a is a variable. However as we shall see, we can still use the results obtained in (1) and (5) in the overall system dimensioning problem. But before we proceed, we first formalize the notion of data aggregation in the next subsection.

5.1. A model for data aggregation

The most commonly used model for data aggregation [2,4] assumes that a cluster head collects the packets from all the nodes in its cluster, and after processing and fusion produces a single packet. It is further assumed that irrespective of the number of nodes in the cluster, the size of this aggregated packet is fixed, i.e., it does not depend on the number of packets that were aggregated during data fusion. While this approach keeps

things tractable, the actual extent of aggregation that is possible is determined by the application. In most applications it may not be possible to fuse data from an arbitrary number of nodes into a single packet of fixed sized. In general we expect the size of the aggregated data packet to increase with an increase in the number of input packets.

We propose a simple model for data aggregation that accounts for the above observation. Consider a cluster with a single cluster head node and x sensor nodes. We assume that the node density is constant, and hence the number of nodes in each cluster, x , is proportional to the area of the cluster. During each data gathering cycle the cluster head receives x packets from the nodes in its cluster, performs data aggregation and produces $\chi(x)$ packets (of the same length). Thus the number of the output packets is a function of the number of the input packets. We use the following model for $\chi(x)$, the number of packets in the aggregated output,

$$\chi(x) = mx + c. \quad (13)$$

In this model c corresponds to the overhead of aggregation, while m is the compression ratio.

Note that $m \leq 1$ because in general the data aggregation process does not increase the per packet payload of the input. We note that this model captures the following aggregation models depending on the values of m and c :

- If $m = 0, c > 0$ then (13) corresponds to the case when any number of packets can be compressed into a single packet of fixed length. This is the model used in [2,4,5]. This models those applications where we want updates of the type *min*, *max* (e.g. temperature), *sum* (e.g. event count), and *yes-no* (e.g. intrusion detection and other 0–1 event detection sensor networks).
- If $m < 1, c > 0$ then (13) corresponds to the case when there is a fixed compression ratio that can be achieved. This could be used to model scenarios in which the data bytes of all the received packets can be compressed by a factor of m . It could also be used to model the scenario in which the cluster head node uses its own address in the aggregated packet to reduce the redundant addressing overheads.

- If $m = 1$ then (13) corresponds to the case when there is no data aggregation. Although clustering benefits from data aggregation, data aggregation may not be the only reason for using clusters. For sensor networks with a large number of nodes, scalability is an important issue. Clustering makes the system scalable. Instead of having a centralized control over thousands of nodes, or having a distributed protocol that operates over thousands of nodes, it is better to organize nodes into smaller clusters, and assign the responsibility of MAC and routing in each cluster to a single cluster head node.

We note that for large clusters (large x), it may not always be possible to sustain the same compression ratio of m , since the correlation between the measured data in a large cluster may not be sufficient for a compression ratio of m . In such cases we require a more elaborate model in which $\chi(x)$ is not linear in x , but a more general function. Such a function can only be defined by knowing the exact correlation structure of the phenomenon that is being sensed. However the model in (13) fits well for several phenomena of practical interest. In this model, m and c are the inputs from the application and they serve as an entry point for the application in the overall network design problem. We believe that the network should be designed by taking into account the extent of data aggregation that is possible when using clustering. The assumption that *irrespective of the size of the cluster, all the packets can be aggregated into a single packet of fixed length* is extremely restrictive and is not a good model for most sensor networks. With (13) as our model for data aggregation, we now address the problem of determining the optimum number of cluster heads, the required battery energy of nodes and the optimum communication mode (whether to use multi-hop or single hop within a cluster) for a general sensor network.

5.2. Overall system design problem

In Section 4 we studied the scenario in which there was a single cluster head located at the

center of a circular region. The motivation behind studying this seemingly over-simplified model was to use it as a building block in the overall network design problem. Consider a circular region of radius A over which n_0 sensor nodes are randomly and uniformly distributed. The number of sensor nodes n_0 is determined by the application requirements and is assumed to be fixed. A remote base station is located at a distance d from the center of the region. We assume that m and c in (13) have been provided by the application. The problem we wish to address is as follows:

1. What is the optimum number of cluster heads, n_1 ?
2. How should we dimension the battery energies of both types of nodes to ensure at least T data gathering cycles?
3. What is the optimum mode for communication between the sensor nodes and the cluster heads, single hop or multi-hop?

Since the base station is located outside the region, the communication between the cluster heads and the base station is single hop.

We assume a propagation loss constant of k for communication within a cluster, and k' for communication between the cluster heads and the base station. Since the cluster head to base station communication is long range, it is likely that $k' > k$. The exact values of k and k' depend on the environment in which the network operates. The authors in [2] assume $k = 2$ (which need not be the case in general) for communication within each cluster, and use single hop communication between the sensor nodes and the cluster heads. They note that for the system parameters that they investigate, multi-hop mode results in more energy expenditure than single hop mode, because the energy spent in transmitter/receiver electronics (l) is comparable to the energy spent in the power amplifier (μx^k). However when we consider a general sensor network that may be deployed over a large region (large x), the μx^k term may dominate the l term to such an extent that using multi-hop mode may be more energy-efficient than single hop mode. Hence it is necessary to compare both single

hop and multi-hop communication modes for the most general network settings.

We showed in Section 4 through (1) and (5) that when $k = 2$, single hop mode is more energy efficient for each individual cluster. However when $k > 2$, the choice between single hop and multi-hop modes depends on the radius of the cluster, a . In a network design problem, the radius of a cluster, a is itself a variable. Hence we formulate two optimization problems. The first problem assumes single hop mode within the clusters while the second problem assumes multi-hop mode within the clusters. We find the optimum choice of system parameters for both the settings and then compare the two solutions to decide which solution is better.

There are two approaches to designing clustered sensor networks. In LEACH [2], the cluster head nodes are selected from among the sensor nodes, and then the cluster heads are rotated periodically for load balancing. While this solution leads to a more uniform energy drainage pattern in the network, it has the disadvantage of adding extra complexity to all the nodes. In this scheme every node has complex hardware and software to act as a cluster head. This involves co-ordinating MAC, routing, data fusion and performing long range transmissions to the distant base station. Besides, energy is also spent on periodic cluster head re-election protocol.

Another approach that has been taken in [8] is that of using heterogeneous nodes. The authors use two types of nodes; type 0 nodes and type 1 nodes. The type 0 nodes are the sensor nodes that perform the job of sensing and sending the sensed data to the cluster heads. The type 1 nodes serve as the cluster heads. They are provided with more battery energy and extra hardware and software complexity. Thus there is no need for a cluster head election protocol, since the cluster head nodes are predetermined. The right objective function to minimize in such a scenario is not the overall energy expenditure, but the overall cost of the network (which takes into account the hardware complexity as well as the battery energy of the nodes). We take this approach, i.e., we assume two types of nodes and minimize the overall network cost.

Note that we are not re-solving the same problems that were solved in [2,4,8]. Instead, we are solving those problems with two important generalizations that are typical of real life sensor networks, and that were not accounted for in the above studies:

1. A fair comparison of multi-hop and single hop mode.
2. A more general model for data aggregation, namely $\chi(x)$.

5.3. Problem formulation

We assume that n_1 type 1 nodes are randomly and uniformly distributed over the region in addition to the n_0 type 0 nodes. Let E_0 be the battery energy of type 0 nodes, and E_1 be the battery energy of type 1 nodes. As in [8], we model the cost of a type i node as follows:

$$C_i = \alpha_i + \beta E_i,$$

where α_i is the hardware cost of the node, while the second term accounts for the battery cost of the node. The constants α_i and β depend on the manufacturing process. We could also use β to model the weight and/or size of the battery. In many commercial sensor nodes, the bulk of the weight and volume of the node is occupied by the battery. If one of the constraints of sensor node design is to limit the weight of the sensor node, then β could be used to model the weight of the node. The higher the required battery energy, the larger the weight of the battery, and hence the larger the weight of the node. We assume that the number of sensor nodes n_0 is fixed (depending on the application requirements, see Section 3). We would like to determine n_1 , E_0 and E_1 so as to minimize the overall network cost which is given by

$$f(n_1, E_0, E_1) = n_0(\alpha_0 + \beta E_0) + n_1(\alpha_1 + \beta E_1). \quad (14)$$

Depending on whether we use single hop or multi-hop communication within the clusters, we obtain different cost functions. Let $f_s(n_1, E_0, E_1)$ denote the cost of the single hop sensor network and $f_m(n_1, E_0, E_1)$ denote the cost of the multi-hop

sensor network. Our plan is to obtain parameters that minimize the cost of both the sensor networks and then compare these minimized costs to determine which scheme is better.

Since there are n_0 type 0 nodes and n_1 cluster heads, the average number of type 0 nodes in each cluster is n_0/n_1 . At each cluster head, during every data gathering cycle, energy is spent on receiving n_0/n_1 packets from the sensor nodes, aggregating them into $\chi(n_0/n_1)$ packets, and transmitting the aggregated packets to the distant base station. In order to sustain T data gathering cycles, the battery energy of a type 1 node should be

$$E_1 = T \left(\frac{n_0}{n_1} (l + E_f) + \chi \left(\frac{n_0}{n_1} \right) (l' + \mu' d^{k'}) \right), \quad (15)$$

where E_f is the computational energy spent on fusion of each packet. As discussed in Section 4, l' and μ' are per packet quantities. Hence $l' + \mu' d^{k'}$ is the energy spent on transmitting a packet from the cluster head to the base station. Note that for a fixed n_1 , E_1 is fixed irrespective of whether the sensor nodes use single hop or multi-hop communication to reach the cluster head.

5.4. Single hop mode

Since the area of the region is πA^2 , we can approximate each cluster to be a circular region of area $\pi A^2/n_1$, i.e., of radius $A/\sqrt{n_1}$. When single hopping is used within the cluster, using (1), the required battery energy of a type 0 node E_0^s is

$$E_0^s = T \left(l + \mu \left(\frac{A}{\sqrt{n_1}} \right)^k \right) = T \left(l + \frac{\mu A^k}{n_1^{k/2}} \right). \quad (16)$$

Hence using (14)–(16) we obtain $f_s(n_1, E_0, E_1)$ as follows:

$$\begin{aligned} f_s(n_1) &= n_0 \alpha_0 + n_0 \beta T \left(l + \frac{\mu A^k}{n_1^{k/2}} \right) + n_1 \alpha_1 \\ &\quad + n_1 \beta T \left(\frac{n_0}{n_1} (l + E_f) \right. \\ &\quad \left. + \chi \left(\frac{n_0}{n_1} \right) (l' + \mu' d^{k'}) \right) \end{aligned}$$

$$\begin{aligned} \Rightarrow f_s(n_1) &= n_0 (\alpha_0 + 2\beta T l + \beta T E_f) + \frac{n_0 \beta T \mu A^k}{n_1^{k/2}} \\ &\quad + n_1 \alpha_1 + \beta T (l' + \mu' d^{k'}) n_1 \chi(n_0/n_1) \end{aligned} \quad (17)$$

$$\begin{aligned} &= A_s + \frac{B_s}{n_1^{k/2}} + C n_1 + D n_1 \chi(n_0/n_1) \\ &= A_s + \frac{B_s}{n_1^{k/2}} + (C + Dc) n_1 + D m n_0. \end{aligned} \quad (18)$$

Thus $f_s(\cdot)$ is a function of just one variable n_1 (n_0 is fixed). Constants A_s , B_s , C and D have been introduced for ease of notation. The optimum number of cluster heads for single hop communication, $n_1 = N_s$, is obtained by minimizing (18).

$$\begin{aligned} \frac{d}{dn_1} f_s(n_1) &= \frac{-k B_s}{2 n_1^{(k+2)/2}} + C + Dc = 0 \\ \Rightarrow N_s &= \left(\frac{k B_s}{2(C + Dc)} \right)^{2/(k+2)} \\ \Rightarrow N_s &= \left(\frac{k n_0 \beta T \mu A^k}{2(\alpha_1 + c \beta T (l' + \mu' d^{k'}))} \right)^{2/(k+2)}. \end{aligned} \quad (19)$$

The second derivative of $f_s(n_1)$ is always positive and hence the above solution is a global minimum. The cost corresponding to the above solution is $f_s(N_s)$.

5.5. Multi-hop mode

Let R be the radius of communication for the multi-hop mode. Since we can approximate each cluster to be a circular region of radius $A/\sqrt{n_1}$, using (5) the required battery energy for a type 0 node as a function of R is

$$\begin{aligned} E_0^m(R) &= T \left(\frac{2l \left(\frac{A}{\sqrt{n_1}} \right)^2}{R^2} - l + \mu R^{k-2} \left(\frac{A}{\sqrt{n_1}} \right)^2 \right) \\ &= T \left(\frac{2l A^2}{n_1 R^2} - l + \frac{\mu R^{k-2} A^2}{n_1} \right) \\ &= T \left(\frac{A^2 (2l + \mu R^k)}{n_1 R^2} - l \right). \end{aligned} \quad (20)$$

Note that we implicitly assumed that R , the thickness of each ring in a cluster, is less than the average radius of the cluster. Hence in case of multi-hop communication we have the following additional constraint:

$$R \leq \frac{A}{\sqrt{n_1}} \Rightarrow n_1 R^2 \leq A^2. \quad (21)$$

We also observed in Section 4 that for multi-hop communication to be possible, the communication radius should be sufficiently large to ensure connectivity with high probability. If it is required to have connectivity with a probability of at least $1 - \epsilon$, the corresponding minimum communication radius can be determined as in (2), and we require

$$R \geq r = A \sqrt{\frac{1}{n_0} \log \left(\frac{n_0}{\epsilon} \right)}. \quad (22)$$

We also require that the communication radius R be smaller than A . Otherwise the scheme is the same as the single hop mode with a single cluster head.

$$R \leq A. \quad (23)$$

Since the radius of communication R for multi-hop communication is another variable at our disposal, we obtain $f_m(\cdot)$ as a function of R and n_1 as follows:

$$\begin{aligned} f_m(R, n_1) &= n_0(\alpha_0 + \beta TE_f) + \frac{n_0 \beta T A^2 (2l + \mu R^k)}{n_1 R^2} \\ &\quad + n_1 \alpha_1 + \beta T (l' + \mu' d^k) n_1 \chi(n_0/n_1) \quad (24) \\ &= A_m + \frac{B_m (2l + \mu R^k)}{n_1 R^2} + C n_1 + D n_1 \chi(n_0/n_1), \end{aligned} \quad (25)$$

where A_m and B_m are appropriately defined constants.

To minimize the cost of multi-hop network, we note that under the assumption (13), i.e., $\chi(x) = mx + c$, the cost function in (25) has the following form:

$$\begin{aligned} f_m(R, n_1) &= \left(\frac{B_m}{n_1} \right) \left(\frac{2l + \mu R^k}{R^2} \right) \\ &\quad + \left(A_m + C n_1 + D n_1 \left(m \frac{n_0}{n_1} + c \right) \right) \\ &= \frac{v(R)}{n_1} + \gamma n_1 + \delta, \end{aligned} \quad (26)$$

where

$$v(R) = \frac{n_0 \beta T A^2 (2l + \mu R^k)}{R^2}.$$

We would like to minimize the cost function in (26) with (21)–(23) as constraints. This is a standard non-linear optimization problem that can be solved using the Karush–Kuhn–Tucker (KKT) theorem. Let $\bar{y} = [R, n_1]$. Then the optimization problem can be formulated as follows:

$$\begin{aligned} &\text{minimize} && f_m(\bar{y}) \\ &\text{subject to} && g_1(\bar{y}) = n_1 R^2 - A^2 \leq 0, \\ & && g_2(\bar{y}) = r - R \leq 0, \\ & && g_3(\bar{y}) = R - A \leq 0. \end{aligned}$$

Note that when the constraint $g_1(\bar{y})$ is active, i.e., when $g_1(\bar{y}) = 0$, we have $n_1 R^2 = A^2$. This effectively means that the thickness of each ring, R , is equal to the radius of the cluster $A/\sqrt{n_1}$, i.e., the mode of communication is effectively single hop. It is easy to verify that when this happens $f_m(\cdot)$ in (24) reduces to $f_s(\cdot)$ in (17). Thus single hop is a special case of multi-hop with $n_1 R^2 = A^2$. Similarly when the constraint $g_3(\bar{y})$ is active, i.e., when $g_3(\bar{y}) = 0$, we have $R = A$. This is another special case of single hop in which there is just one cluster. Minimizing the same function under an additional constraint of $g_2(\bar{y}) = r - R \leq 0$ will lead to a cost function which can only be larger than the unconstrained minimization of the same function in (17). Hence we conclude that when the constraints $g_1(\bar{y})$ or $g_3(\bar{y})$ become active, single hop mode has a lower cost, and therefore having already solved the single hop problem in the previous subsection, we need not solve the multi-hop problem for these two cases.

If the constraints $g_1(\bar{y})$ and $g_3(\bar{y})$ are inactive, i.e., $g_1(\bar{y}) < 0$ and $g_3(\bar{y}) < 0$, we can simply minimize the cost function in (26) with $g_2(\bar{y})$ as the only constraint. We should of course verify that the solution thus obtained is indeed feasible, i.e., $g_1(\bar{y}) < 0$ and $g_3(\bar{y}) < 0$. Let $\nabla f(\bar{y})$ denote the gradient vector of function $f(\bar{y})$:

$$\begin{aligned} \nabla f(\bar{y}) &= \begin{bmatrix} v'(R)/n_1 \\ (-v(R)/n_1^2) + \gamma \end{bmatrix}, \\ \nabla g_1(\bar{y}) &= \begin{bmatrix} 2n_1 R \\ R^2 \end{bmatrix}, \\ \nabla g_2(\bar{y}) &= \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \nabla g_3(\bar{y}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \end{aligned} \quad (27)$$

Using the KKT theorem, the solution to the optimization problem is

$$\nabla f(\bar{y}) + \mu_1 \nabla g_1(\bar{y}) + \mu_2 \nabla g_2(\bar{y}) + \mu_3 \nabla g_3(\bar{y}) = \mathbf{0} \quad (28)$$

with

$$\begin{aligned} \mu_1 \geq 0, \quad \mu_2 \geq 0, \\ \mu_1 g_1(\bar{y}) + \mu_2 g_2(\bar{y}) + \mu_3 g_3(\bar{y}) = 0, \end{aligned} \quad (29)$$

where μ_1 , μ_2 and μ_3 are the constants of the KKT theorem.

The case of $\mu_1 \neq 0$, i.e., $g_1(\bar{y}) = 0$ as well as the case of $\mu_3 \neq 0$, i.e., $g_3(\bar{y}) = 0$ correspond to single hop optimization problem (which we have already solved in Section 5.4) as already pointed out. So we only look at the case when $\mu_1 = \mu_3 = 0$. This along with (29) gives two solutions; $\mu_2 = 0$ and $g_2(\bar{y}) = 0$, i.e., $R = r$.

The case of $\mu_2 = 0$ (along with $\mu_1 = \mu_3 = 0$) corresponds to the unconstrained optimization and results in $\nabla f(\bar{y}) = 0$.

$$\frac{\partial}{\partial R} v(R) = 0 \Rightarrow \frac{\partial}{\partial R} \left(\frac{2l + R^k}{R^2} \right) = 0.$$

We have already seen in Section 4, that the solution to the above equation is \hat{R} and is given by (8):

$$\hat{R} = \left(\frac{4l}{\mu(k-2)} \right)^{1/k}. \quad (30)$$

Also

$$\begin{aligned} \frac{\partial}{\partial n_1} f_m(R, n_1) &= -\frac{v(R)}{n_1^2} + \gamma \\ &= \frac{-B_m(2l + \mu R^k)}{R^2 n_1^2} + C + Dc. \end{aligned} \quad (31)$$

Setting the above derivative to zero and with $R = \hat{R}$, we obtain the optimum number of cluster heads $n_1 = N_m(\hat{R})$ as follows:

$$\begin{aligned} N_m(\hat{R}) &= \left(\frac{B_m(2l + \mu \hat{R}^k)}{\hat{R}^2(C + Dc)} \right)^{1/2} \\ &= \left(\frac{n_0 \beta T A^2 (2l + \mu \hat{R}^k)}{\hat{R}^2 (\alpha_1 + c \beta T (l' + \mu' d^{k'}))} \right)^{1/2}. \end{aligned} \quad (32)$$

Note that the feasibility of the above solution $[\hat{R}, N_m(\hat{R})]$ needs to be verified by checking for

$$N_m(\hat{R}) < \frac{A^2}{\hat{R}^2} \quad (33)$$

and

$$A > \hat{R} > r. \quad (34)$$

The verification of the feasibility of the solution can only be done on a per case basis depending on the system parameters. The other possible solution corresponds to $R = r$ with $\mu_1 = \mu_3 = 0$. We obtain the optimum number of cluster heads $n_1 = N_m(r)$ as follows:

$$\begin{aligned} \nabla f(\bar{y}) + \mu_2 \nabla g_2(\bar{y}) = \mathbf{0} &\Rightarrow -\frac{v(r)}{n_1^2} + \gamma + \mu_2(0) = 0 \\ \Rightarrow N_m(r) &= \left(\frac{n_0 \beta T A^2 (2l + \mu r^k)}{r^2 (\alpha_1 + c \beta T (l' + \mu' d^{k'}))} \right)^{1/2}. \end{aligned} \quad (35)$$

For the above solution to be feasible we must verify that

$$N_m(r) < \frac{A^2}{r^2} \quad (36)$$

and $\mu_2 \geq 0$. Since $\nabla f(\bar{y}) + \mu_2 \nabla g_2(\bar{y}) = \mathbf{0}$, from (27) we require

$$\mu_2 = \frac{v'(r)}{n_1} \geq 0 \Rightarrow v'(r) = -\frac{4l}{r^3} + (k-2)r^{k-3} \geq 0. \quad (37)$$

Note that $r < A$ is always true, since a communication radius of A trivially ensures connectivity. If the above solution is feasible, the corresponding cost is $f_m(r, N_m(r))$.

We can also prove that $f_m(\hat{R}, N_m(\hat{R}))$ and $f_m(r, N_m(r))$ correspond to local minimum and not local maximum when the feasibility conditions are satisfied. For this, we use the second order sufficiency test of the KKT theorem. Let $\mathbf{F}(\bar{y})$ be the Hessian matrix corresponding to function $f_m(R, n_1)$, and $\mathbf{G}_i(\bar{y})$ be the Hessian matrix corresponding to $g_i(R, n_1)$. Let

$$\begin{aligned} \mathbf{L}(y^*, \mu^*) &= \mathbf{F}(y^*) + \mu_1 \mathbf{G}_1(y^*) + \mu_2 \mathbf{G}_2(y^*) \\ &\quad + \mu_3 \mathbf{G}_3(y^*), \end{aligned} \quad (38)$$

where Hessian is defined as follows:

$$\mathbf{F}(\bar{\mathbf{y}}) = \begin{bmatrix} \frac{\partial^2 f_m}{\partial R^2}(\bar{\mathbf{y}}) & \frac{\partial^2 f_m}{\partial n_1 \partial R}(\bar{\mathbf{y}}) \\ \frac{\partial^2 f_m}{\partial R \partial n_1}(\bar{\mathbf{y}}) & \frac{\partial^2 f_m}{\partial n_1^2}(\bar{\mathbf{y}}) \end{bmatrix}.$$

Since $\mu_1 = \mu_3 = 0$ for both the solutions corresponding to $R = \hat{R}$ and $R = r$, the second order sufficient condition for local minimization (see [10] for details) is

$$u^T \mathbf{L}(\mathbf{y}^*, \mu^*) u > 0 \quad \forall u: \nabla g_2(\mathbf{y}^*) \cdot u = 0.$$

From (27) the tangent space which corresponds to $u: \nabla g_2(\mathbf{y}^*) \cdot u = 0$ is the space of all the vectors of the form $[0, \theta]$. Hence we require that $\mathbf{L}(\mathbf{y}^*, \mu^*)$ be positive definite for all the vectors of the form $[0, \theta]$. We have

$$\mathbf{F}(R, n_1) = \begin{bmatrix} v''(R)/n_1 & -v'(R)/n_1^2 \\ -v'(R)/n_1^2 & 2v(R)/n_1^3 \end{bmatrix}. \quad (39)$$

For both the solutions $R = \hat{R}$ and $R = r$, we have $\mu_1 = 0$ and $\mathbf{G}_2(\bar{\mathbf{y}}) = \mathbf{0}$. Hence proving that $\mathbf{L}(\mathbf{y}^*, \mu^*)$ in (38) is positive definite for vectors of the form $[0, \theta]$ is equivalent to proving that $\mathbf{F}(\mathbf{y}^*)$, i.e., (39) is positive definite for vectors of the form $[0, \theta]$. This in turn is equivalent to proving that

$$\theta^2 \frac{2v(R)}{n_1^3} > 0 \quad \text{for } R = \hat{R}, R = r \text{ and } \forall \theta.$$

Since $v(R) > 0$ for all R we conclude that when the solutions $R = \hat{R}$ and $R = r$ are feasible, they minimize the cost function.

5.6. Summary

Thus in order to determine the optimum number of cluster heads, the optimum communication mode and the optimum radius of communication (if multi-hop communication is used), we must determine \hat{R} , r , N_s , $N_m(\hat{R})$ and $N_m(r)$, verify the feasibility conditions for the latter two solutions, determine the corresponding costs for all the feasible solutions, and pick the solution that has the lowest cost.

We know that $f_m(\hat{R}, N_m(\hat{R}))$ corresponds to the unconstrained minimization, and single hopping is a special case of multi-hopping as seen in Section

5.5. Hence if $[\hat{R}, N_m(\hat{R})]$ is feasible, then it is the desired minimum cost solution. However if this solution is not feasible, we must determine $f_s(N_s)$ and $f_m(r, N_m(r))$. If $[r, N_m(r)]$ is also not feasible, then the only solution is N_s , i.e., single hop mode. However if $[r, N_m(r)]$ is feasible, then we must compare the costs $f_s(N_s)$ and $f_m(r, N_m(r))$ and choose the solution with a lower cost.

Note that the solutions for N_s and N_m in (19) and (32) have an altogether different form depending on the value of k . We further note that these expressions depend only on c and do not depend on m . It can also be shown that the difference in the overall costs of single hop and multi-hop solutions is also independent of m . The reason is that due to the fixed compression ratio m , out of the total data that is gathered during each cycle, a fraction m of that data has to be sent to the base station irrespective of the number of cluster heads and the mode of communication. However m comes into picture when determining the required battery energy of a type 1 node (15). The required battery energy of a type 0 node can be determined from (16) or (20) depending on the choice of communication mode.

The above optimization problem can also be solved in the context of 3-D and 1-D clustered sensor networks by using expressions for \hat{R}_{3D} and \hat{R}_{1D} in (11) and (12) respectively. If the phenomenon to be sensed is governed by a different data aggregation model $\chi(x)$, we can use a similar approach to solve the general optimization problem.

Thus we see that there is no single answer to the question ‘‘which is the best communication mode, single hop or multi-hop?’’. The answer depends on various system parameters such as the radio constants of the surrounding environment and the transceiver (l , l' , μ , k , μ' and k'), the size and the dimensions of the region (A , 1-D, 2-D or 3-D), the distance of the base station from the region (d), the production costs of the nodes (α_1 and β), the required number of sensor nodes as dictated by the application (n_0), the desired lifetime of the network (T), the compressibility of data which in turn is governed by the application (m and c), the computational energy spent on data aggregation (E_f) and the desired probability of connectedness (ϵ , if multi-hop communication is to be used).

6. A hybrid communication mode

In this section we propose a hybrid mode for communication between the sensor nodes and the cluster heads. In the previous section we noted that in single hop mode the sensor nodes which are farthest from the cluster head have the highest energy drainage. By assuming power control functionality in single hop mode, it is possible for the sensor nodes which are closer to the cluster head to transmit at lower power. In multi-hop mode the sensor nodes that are closest to the cluster head have the highest energy drainage due to packet relaying. We propose a scheme in which the sensor nodes alternate between single hop mode and multi-hop mode periodically. When single hop mode is used (along with power control at the nodes) the nodes near the cluster head are relieved of their relaying burden, and when multi-hop mode is used the nodes which are farthest from the cluster head are relieved of their burden of long range transmissions to the cluster head. Thus by alternating between the two modes of communication it is possible to obtain a more uniform load distribution. This is a form of role rotation. A simple way to implement a scheme like this would be to have the cluster head co-ordinate the periodic switch-over. The cluster head can broadcast a beacon periodically to all the nodes in its cluster asking them to switch between the two communication modes. The exact fraction of the time for which each of the two modes is sustained can be easily computed as seen below.

In [9], the authors have provided bounds on the lifetime of a sensor network via optimal role assignment. The idea is to use different paths (not necessarily using the nearest node as the next hop node) for relaying of packets, and to determine the fraction of time for which each of the paths should be sustained so as to minimize the overall energy expenditure. As the number of nodes increases, the number of possible routes blows up exponentially. However using the approach of network flows, it is possible to solve the problem in polynomial time. The approach provides an upper bound on the lifetime of the network over all the possible collaborative data gathering strategies. However implementing such a scheme is difficult, since it is

necessary to know the exact locations of all the nodes, and then to co-ordinate all the nodes so that different collaborative strategies are sustained over different periods.

Our scheme is sub-optimal in that it does not take into account all the possible multi-hop paths. Instead we use just two modes of communication; single hopping and multi-hopping (with some optimum communication radius). The nodes alternate between these two modes periodically. *Our scheme is very easy to implement and does not require the exact knowledge of the node locations.* For this scheme we can determine the optimum number of cluster heads and the battery energies. We can easily prove that this hybrid scheme is better than using pure single hop, or pure multi-hop communication.

We use the same notations as in Section 5. Assume that out of the desired lifetime of T cycles, nodes use single hop communication mode for ϕT cycles and multi-hop communication mode for $(1 - \phi)T$ cycles where $0 \leq \phi \leq 1$. Using power control, the energy spent by a node located at a distance of nR from the cluster head during the ϕT cycles of single hop communication is

$$E_s(nR) = \phi T(l + \mu R^k n^k).$$

Similarly, using (4) and (3) and the communication model of $l + \mu x^k$, the energy spent during the $(1 - \phi)T$ cycles of multi-hop communication is

$$E_m(nR) = (1 - \phi)T \left((2l + \mu R^k) \left(\frac{a^2 - n^2 R^2}{R^2(2n - 1)} \right) + l + \mu R^k \right),$$

where we assume that multi-hopping with a radius of R with n_1 cluster head nodes is feasible. If multi-hopping is not feasible, $\phi = 1$, i.e., we use only single hop mode. Hence the total battery energy required is

$$E_0(nR) = E_s(nR) + E_m(nR).$$

Since the battery energy dimensioning is to be done for the worst case energy expenditure, the actual battery energy allocated to the sensor nodes is the maximum value of $E_0(nR)$ over all the values of n . Note that $E_s(nR)$ is a convex increasing function of n while $E_m(nR)$ is a convex decreasing

function of n . Since n is the ring number, it is a measure of the distance from the cluster head. Hence $E_0(nR)$ is a convex function. Therefore it takes its maximum value at either one or both of the endpoints; $nR = R$, $nR = A/\sqrt{n_1}$, where we used the fact that the average radius of a cluster is $A/\sqrt{n_1}$, and this constitutes the farthest ring. This is easier to see in Fig. 2. At the endpoint $nR = R$, i.e., in the first ring, we already have an expression for $E_m(R)$ from (5). Similarly for the last ring, we have an expression for $E_s(A/\sqrt{n_1})$ from (1). We substitute $A/\sqrt{n_1}$ for a in (5) and (1). We also know that for the first ring, $E_s(R) = l + \mu R^k$ and for the last ring, $E_m(A/\sqrt{n_1}) = l + \mu R^k$, since these involve a single transmission over a distance of R . For ease of notation, let

$$e_0 = (l + \mu R^k), \quad (40)$$

$$e_1 = \left((2l + \mu R^k) \left(\frac{A^2}{n_1 R^2} - 1 \right) + (l + \mu R^k) \right), \quad (41)$$

$$e_2 = \left(l + \mu \frac{A^k}{n_1^{k/2}} \right). \quad (42)$$

Hence we obtain the following expression for the required battery energy as a function of ϕ as follows:

$$\begin{aligned} E_0(\phi) &= \max \left\{ E_0(R), E_0 \left(\frac{A}{\sqrt{n_1}} \right) \right\} \\ &= \max \left\{ E_s(R) + E_m(R), E_s \left(\frac{A}{\sqrt{n_1}} \right) + E_m \left(\frac{A}{\sqrt{n_1}} \right) \right\} \\ &= \max \{ \phi T e_0 + (1 - \phi) T e_1, \phi T e_2 + (1 - \phi) T e_0 \} \\ &= T \max \{ -(e_1 - e_0)\phi + e_1, (e_2 - e_0)\phi + e_0 \}. \end{aligned} \quad (43)$$

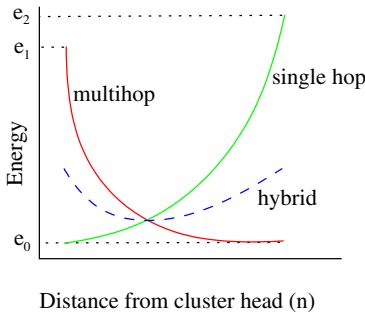


Fig. 2. Hybrid communication mode.

Note that $e_1, e_2 > e_0$ since e_1 and e_2 correspond to the maximum energy expenditure while e_0 corresponds to the minimum energy expenditure for each mode (see Fig. 2). Also note that $\phi = 1$ corresponds to pure single hop mode while $\phi = 0$ corresponds to pure multi-hop mode. As a function of ϕ , $(e_2 - e_0)\phi + e_0$ is linearly increasing, while $-(e_1 - e_0)\phi + e_1$ is linearly decreasing. Hence the max of the two functions is minimized for the value of ϕ at which the two functions become equal. Let ϕ_0 be that value of ϕ .

$$\begin{aligned} (e_2 - e_0)\phi_0 + e_0 &= -(e_1 - e_0)\phi_0 + e_1 \\ \Rightarrow \phi_0 &= \frac{e_1 - e_0}{e_2 + e_1 - 2e_0} \end{aligned} \quad (44)$$

$$\Rightarrow E_0 = E_0(\phi_0) = T \left(\frac{e_1 e_2 - e_0^2}{e_2 + e_1 - 2e_0} \right). \quad (45)$$

Thus $E_0(\phi_0) \leq E_0(1) = E_s$ and $E_0(\phi_0) \leq E_0(0) = E_m$. Thus for the same number of cluster heads (n_1) and the same communication radius (R), the hybrid scheme results in a lower battery energy for type 0 nodes as compared to pure single hop or pure multi-hop modes. Since the energy requirement of the type 1 nodes, i.e., E_1 is not affected by the communication mode within the cluster, the overall cost of the network is also lower for the hybrid mode as compared to pure single hop or pure multi-hop mode. If $f(\bar{y}) \leq g(\bar{y})$, then the minimum value of $f(\bar{y})$ is also less than or equal to the minimum value of $g(\bar{y})$. Hence the hybrid mode is more cost effective than both single hop as well as multi-hop modes.

Having obtained an expression for E_0 , we must now determine the optimum number of cluster heads, n_1 and the radius of communication R for multi-hop communication. For this we substitute the expression for E_0 from (45) using (40)–(42) in the cost function along with E_1 (given by (15)), and then minimize the cost function under (21)–(23) as constraints to determine n_1 and R . The optimization problem can again be solved using the KKT theorem as in Section 5. However unlike Section 5, in this case the equations are much more complicated and hence it is difficult to obtain closed form solutions for n_1 and R . However for a given scenario of interest it is possible to solve the equations numerically. Note that we must verify that the

solution thus obtained is indeed feasible, i.e., multi-hopping is indeed possible. If not, the only feasible solution is to use pure single hop communication. Once n_1 and R have been determined, we can determine ϕ_0 using (44).

In our original model we had assumed a remotely located base station. However we can also consider a case in which the base station is located at the center of the region. In this case, we have the same problem at the hierarchy of the base station and the cluster heads, as we had at the hierarchy of a cluster head and the sensor nodes. In this case the cluster heads that are closer to the base station have to perform short range transmissions while the cluster heads located near the edge of the region have to perform long range transmissions to reach the base station. We could again use the hybrid mode of communication in which the cluster heads alternate between single hop mode and multi-hop mode. Multi-hopping is only at the cluster head level, i.e., cluster head nodes use other cluster head nodes as their intermediate hop nodes. For such a scenario, the expression for E_1 is similar to (43) and we can once again solve the corresponding optimization problem.

7. Case studies

In this section we show how the results that we obtained in Section 5 could serve as guidelines to choose the optimum communication architecture, and the optimum number of cluster heads for a given application. We consider two scenarios and show that for the first scenario single hop mode is optimum while for the second scenario multi-hop mode is optimum. For both the scenarios we use similar values for transceiver and propagation loss parameters as in the simulation study of LEACH in [2]. The parameters are close to the state of the art transceivers that are currently available as was pointed out in [2]. Note that in our case l , l' , μ , μ' and E_f are given on a per packet basis (see Table 1), while in [2] the values are given on a per bit basis. The two scenarios that we consider in this section differ in the radio propagation model for communication within the cluster (μ and k). We also assume that $m = 0$ and $c = 1$ for simplicity (N_s

Table 1
System parameters

No. of type 0 nodes, n_0	10^5
Radius of the region, A	1000 m
Distance from base station, d	3000 m
No. of cycles (lifetime), T	10^4
Length of each packet	525 bytes
Aggregation energy, E_f	0.021 mJ/packet
Connectivity probability, $1 - \epsilon$	0.99
Cluster head to base station:	(per packet)
$l' + \mu'x^k$	0.21 mJ + 5.46x ⁴ pJ

and N_m do not depend on m , only E_1 depends on m). The system parameters given in Table 1 are common to both the scenarios.

7.1. Scenario I

This is the scenario when the propagation loss exponent for communication within the cluster, k , is two. Correspondingly, the energy required to transmit a packet over distance x within the cluster is $l + \mu x^2$. For a 525 byte packet this equals 0.21 mJ + 42x² nJ. When $k = 2$, we have already seen in Section 4.3 that single hop mode is more cost effective than multi-hop mode. With the above system parameters we obtain the optimum number of cluster heads N_s using (19). We plot N_s as a function of α_1/β (x axis is divided by a constant $cT(l' + \mu'd^k)$). Note that the product $\alpha_1/\beta cT(l' + \mu'd^k)$ corresponds to the ratio of the hardware cost of the cluster head to its battery cost. This is because in the expression for E_1 in (15), the term corresponding to $cT(l' + \mu'd^k)$ is large as compared to the first term. Thus depending on the manufacturing cost of the hardware (α_1) and the battery cost factor (β) of type 1 nodes we can determine the required number of cluster heads from Fig. 3. In this figure we note that the optimum number of cluster heads, N_s , is between one and three depending on the ratio α_1/β . With three cluster head nodes, i.e., $N_s = 3$, using (15) and (16) and with $k = 2$, we find that the required battery energy of the cluster heads is about 4.5 MJ and the battery energy of the sensor nodes is about 0.14 kJ. Clearly the cluster head nodes have a higher battery energy requirement. The required battery energy of sensor nodes is close to the typical battery energy of some of the

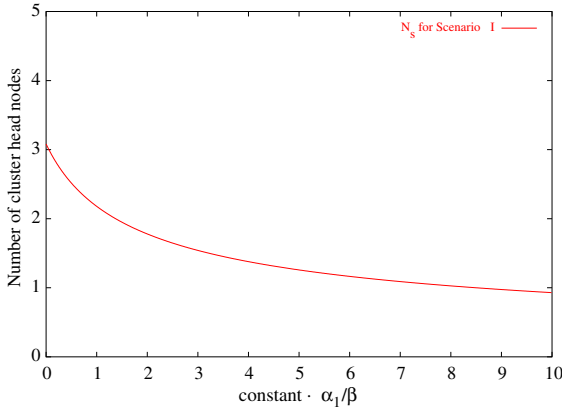


Fig. 3. Scenario I: number of cluster heads as a function of the relative cost of the hardware of a cluster head node, α_1/β .

commercially available sensor nodes. For example, a typical Mote sensor node uses a 3V battery with 560 mAh rating which corresponds to about 6 kJ of energy [11].

Instead of using two types of nodes if we use LEACH, then using the results obtained in [2], we find that the required number of cluster heads for the above settings is about three. However in the case of LEACH each of the 10^5 nodes has the extra hardware and software complexity of a cluster head node. We also find that the required battery energy in the sensor nodes for the above settings when LEACH is used, is about 0.18 kJ.

We understand that having two types of nodes leads to a scheme which is less robust. This is because once the cluster head nodes fail, the system stops functioning. In the case of LEACH the system is more robust because every node is capable of acting as a cluster head, and hence the failure of a few nodes does not seriously affect the working of the system. However it must be noted that this additional robustness comes at an extra cost; the cost of adding the cluster head functionality at each and every node.

7.2. Scenario II

For this scenario we assume that the propagation loss exponent for communication within the cluster, k , is four. As a result, the model for communication within the cluster is the same as

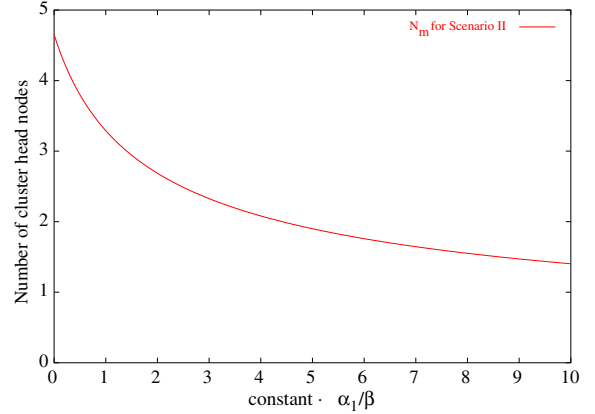


Fig. 4. Scenario II: number of cluster heads as a function of the relative cost of the hardware of a cluster head node, α_1/β .

the model for communication between the cluster heads and the base station. Hence we have $l = l'$, $k = k' = 4$ and $\mu = \mu'$ and the parameters for communication between the cluster heads and the base station (l' , μ' and k') are as shown in Table 1. In this case the surrounding environment is lossy. This is usually the case when sensor nodes are deployed over a region of dense vegetation, buildings or factories where the propagation fall-off is a lot more drastic than free space loss. We find that for this scenario, the multi-hop communication mode turns out to be the optimum choice. In fact we can verify that the conditions in (33) and (34) are hold, and therefore the unconstrained minimization solution is feasible. We therefore obtain $[\hat{R}, N_m(\hat{R})]$ as the optimum solution. For this scenario, using (30) and (22) and with $\mu = \mu' = 5.46$ pJ/m⁴, we obtain $\hat{R} = 94$ m and $r = 13$ m. The dependence of N_m on α_1/β is given in Fig. 4 using (32). We note that depending on the ratio α_1/β , the required number of cluster heads varies between one and five. With three cluster head nodes, i.e., $N_m = 3$, using (15) and (20) we find that the required battery energy for the cluster head nodes is about 4.5 MJ while the required energy for the sensor nodes is about 0.32 kJ.

8. Conclusions

We studied the problem of the design of wireless sensor networks from the point of view of the di-

mensioning of the battery energy of the nodes, the number of cluster heads and the optimum mode of communication between the sensor nodes and the cluster heads. We did a systematic comparative study of single hop and multi-hop modes, and also proposed a new hybrid mode which performs better than both modes. We also proposed a model for data aggregation, and showed how the application can enter the overall system design problem through the data aggregation model $\chi(x)$. We formulated and solved a cost based optimization problem to compare single hop and multi-hop sensor networks. We also formulated a similar problem for the hybrid mode of communication. The results obtained in Sections 5 and 6 could serve as guidelines for the designers of sensor networks to determine the best mode of communication, the optimum number of cluster heads to be used for a given application and, the required battery energies of the nodes.

Acknowledgements

This work was supported in part by a DARPA grant (contract no. MDA 972-02-1-0032) and a grant from the Purdue Research Foundation.

References

- [1] I.F. Akyildiz, W. Su, Y. Sankarsubramaniam, E. Cayirci, Wireless sensor networks: a survey, *Computer Networks* 38 (2002) 393–422.
- [2] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications* 1 (4) (2002) 660–670.
- [3] M. Bhardwaj, T. Garnett, A.P. Chandrakasan, Upper bounds on lifetime of sensor networks, *IEEE International Conference on Communications (ICC'01)*, Helsinki Finland, June 2001.
- [4] S. Bandyopadhyay, E. Coyle, An energy efficient hierarchical clustering algorithm for wireless sensor networks, *IEEE Infocom*, San Francisco, CA, 2003.
- [5] S. Lindsey, C. Raghavendra, Pegasus: power efficient gathering in sensor information systems, *IEEE International Conference on Communications (ICC'01)*, Helsinki Finland, June 2001.
- [6] P. Gupta, P.R. Kumar, Critical power for asymptotic connectivity in wireless networks, in: W.M. McEneaney, G. Yin, Q. Zhang (Eds.), *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, Birkhauser, Boston, MA, 1998, pp. 547–566.
- [7] T.S. Rappaport, *Wireless Communication*, Prentice-Hall, Englewood Cliffs, NJ, 1996.
- [8] V. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, N. Shroff, A minimum cost surveillance sensor network with a lifetime constraint, submitted March 2003. Available from <<http://web.ics.purdue.edu/~mhatre/lifetime.pdf>>.
- [9] M. Bhardwaj, A.P. Chandrakasan, Bounding the lifetime of sensor networks via optimal role assignments, *IEEE Infocom*, New York, 2002.
- [10] E. Chong, S. Zak, *An Introduction to Optimization*, second ed., Wiley, New York, 2001.
- [11] J. Hill, TinyOS—communication and computation at the extremes, 9th International Conference on ASPLOS, Cambridge, MA, USA, November 12–15, 2000. Available form <http://webs.cs.berkeley.edu/tos/presentations/ASPLOS_2000.ppt>.



Vivek Mhatre graduated with a B.Tech degree in Electrical Engineering from the Indian Institute of Technology (IIT) Bombay, India in August 2000. He is currently working towards the Ph.D. degree at the School of Electrical and Computer Engineering at Purdue University, USA. His research interests include wireless sensor networks and ad hoc networks.



Catherine Rosenberg has worked in several countries including USA, UK, Canada, France and India. In particular, she worked for Nortel Networks in the UK, AT&T Bell Laboratories in the USA, Alcatel in France and taught at Ecole Polytechnique of Montreal (Canada). Dr. Rosenberg is currently Professor in the School of Electrical and Computer Engineering at Purdue University. She is also the Director of the university-wide Center for Wireless Systems and Applications at Purdue University. Dr. Rosenberg is an Associate Editor for *IEEE Transactions on Mobile Computing, Telecommunication Systems*, and *IEEE Communications Surveys*. She has been, and is involved in many conferences including *IEEE INFOCOM*, *International Teletraffic Congress (ITC)*, *IEEE International Conference on Communications (ICC)*, and *IEEE Mobicom*. Her research interests are in all the aspects of networking including wireless, peer-to-peer, security, and traffic engineering.