Chapter 1

ENERGY AND COST OPTIMIZATIONS IN WIRELESS SENSOR NETWORKS: A SURVEY

Vivek Mhatre
School of Electrical and Computer Engineering,
Purdue University,
West Lafayette, IN 47907-1285, USA
Phone: (765)-494-0626
mhatre@ecn.purdue.edu

Catherine Rosenberg
Department of Electrical and Computer Engineering,
University of Waterloo,
Waterloo, Ontario, N2L 3G1, Canada
Phone: 519-888-4016, Fax: 519-886-0619
Corresponding author
cath@ece.uwaterloo.ca

Abstract We present a survey of some of the recent work on energy and cost optimizations in wireless sensor networks. Sensor nodes are characterized by severe energy budget due to limited battery life. We focus on two main problem areas, namely routing and design. In sensor networks in which the nodes use multi-hop communication, routing is a major issue. The routing problem in the context of sensor network retains some of the features of the routing problem in ad-hoc networks, but also has some specific characteristics to it, in particular with respect to data-aggregation, addressing, and the many-to-one paradigm (each sensor node wanting to send the collected data to a single base-station). We first discuss the work done on energy efficient routing, and the corresponding optimization problems for maximizing the lifetime of the network. We then discuss some of the optimization problems in the design and dimensioning of sensor networks. Since most potential applications envisioned for sensor networks require high node density, node heterogeneity and hierarchical clustering could be used for better scalability of the protocols. We discuss the results obtained on energy and cost minimization problems in the context of such clustered sensor networks.

1. Introduction

In the past few years, the field of wireless sensor networks has become a key area of research. Sensor networks find applications in several military as well as civilian domains. Sensor networks along with the widespread Internet enable a user to remotely monitor a phenomenon of interest (see Fig. 1.1). See Akyildiz et al. (2002) for a detailed description of potential sensor network applications. Due to the ad-hoc nature of sensor networks and severe battery energy limitations, energy efficient protocols are required at all the layers of the protocol stack. A nice overview of the recent results on sensor network specific optimizations at the different layers of the protocol stack can be found in Akyildiz et al. (2002). However, in this paper we survey in greater details two important problems in sensor networks, namely energy efficient routing and cost efficient network design. The field of wireless sensor network is receiving a lot of attention, and is evolving very fast. It is difficult to provide a comprehensive survey about a field which is not fully mature yet. Hence this paper should be seen more as a snapshot of the state of the art for the above two issues.

Since a sensor network is deployed with an objective of gathering information, for a given initial battery energy, it is desired that the network continues to function and provide data updates for as long as possible. This is referred to as the maximum lifetime problem in sensor networks. During each data gathering phase, nodes spend a part of their battery energy on transmitting, receiving and relaying packets. Hence the routing algorithm should be designed to maximize the time until the first battery expires, or a fraction of the nodes have their batteries expired. In certain low bandwidth sensor networks, besides the battery energy, the channel bandwidth presents itself as another constraint, and the routing problem has to take this into account. While it is easy to show that such an energy efficient routing problem reduces to a linear programming problem, the real challenge lies in devising lightweight and efficient distributed algorithms for solving it.

The problem of cost efficient network design is mainly a problem in the context of clustered networks. In such networks, nodes are organized into clusters with a single cluster head node per cluster. The sensor nodes send their measured data to their closest cluster head node. The cluster head nodes aggregate the received data, and then send it to the base station. The cluster head nodes could either be identical to

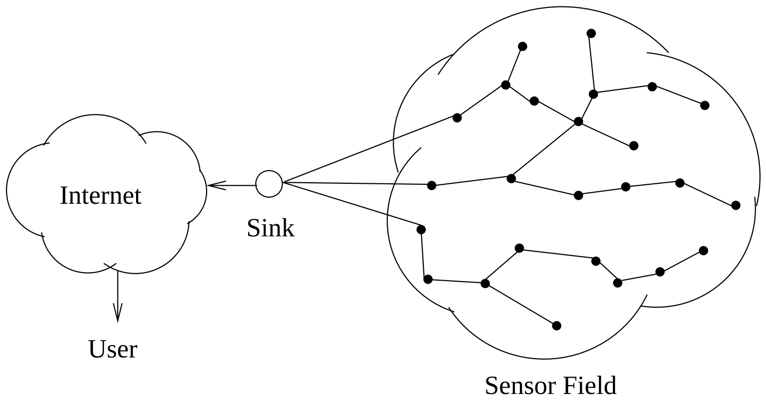

Figure 1.1. A typical sensor network topology.

the sensor nodes (a homogeneous network), or they could be equipped with better hardware and more battery energy than the sensor nodes (a heterogeneous network). In either case, a cost function can be associated with the hardware and the battery cost of each node. From a network designer's perspective the issue is designing the network in such a way that the overall cost of the network is minimized while guaranteeing the desired network lifetime.

This paper is organized as follows. We first provide a brief overview of wireless sensor networks and some of their salient features in section 1.2. In section 1.3, we present a survey of some of the important papers on routing optimizations in sensor networks. Section 1.4 contains a survey of the work on design optimizations in sensor networks. Finally, we conclude in section 1.5.

2. Wireless sensor networks: a brief overview

The purpose of deploying a sensor network is to monitor an area for an event of interest. The advent of affordable wireless technology has led to the vision of empowering small monitoring devices with a wireless network interface that can be used to communicate with other nodes. We discuss some of the salient features of wireless sensor networks in this section.

One of the most important salient features of a sensor network is that the application for which the network is to be used, has a big impact on the design and dimensioning of the network. This is unlike the current Internet where the application has to be designed to work well over the given network. The Internet delivers packets using a best effort service, and so the applications cannot be given any bandwidth or delay guarantees (unless some sophisticated tools such as MPLS are used). Thus the designers of new applications have to work within the framework of the current Internet. However for sensor network applications, it is possible to design and dimension a network in such a way that it caters to the specific requirements of the application. However, the range of sensor network applications is vast. At one end of the spectrum there are applications that require periodic data updates from the network, e.g., temperature monitoring and control in buildings. At the other end of the spectrum are applications in which the network is idle for long periods of time, but bursts into activity as soon as the event of interest occurs, e.g., forest fire detection. In the former case, the traffic is more or less uniform, and there is scope for in-network aggregation of data, while in the latter case, the traffic is bursty, delay-sensitive, and there is no scope for in-network data aggregation. From a designer's viewpoint, the issues involved in designing and dimensioning these two types of networks are altogether different. Hence we classify sensor networks into the following two main categories; data gathering sensor networks and event detection sensor networks. Others have classified the sensor network applications in a more granular way (see Tilak, Abu-Ghazaleh and Heinzelman (2002)). However for the purpose of this survey, this classification suffices.

In data gathering sensor networks, nodes send their measurements periodically to the base station, while in event detection sensor networks, the nodes remain idle most of the time, and spring to activity only when the phenomenon of interest occurs. Most of the work that we describe in this survey paper is about data gathering sensor networks, because routing is an important problem in such networks. On the other hand in event detection sensor networks, MAC and sleep-wake synchronization are the key issues.

The base station could either be located remotely outside the region of interest, or it could be located within the region of interest. In the former case, either all or a few nodes have to perform long range transmission of data to the remote base station. In the latter case, nodes could either use multi-hop communication or direct transmission to reach the base station. The location of the base station is application dependent. For example, in the context of remote surveillance of battlefield, the base station is located far away from the region of interest, while in the context of temperature monitoring and control in buildings, the base station is located in the region of interest.

Sensor networks are also characterized by a many-to-one communication paradigm, i.e., most of the nodes in the network send their data to a few sink nodes. This is unlike the ad-hoc network communication paradigm where each node may wish to communicate with any other node in the network.

The sensor networks that are to be deployed for environmental monitoring and surveillance are expected to be deployed over rough terrain, and are likely to have a high failure rate due to cheap hardware. Node failure, high node density and ad-hoc deployment are some other salient features of most sensor networks. When nodes are deployed randomly, there is likely to be correlation between the measurements of nearby nodes, and thus there is a scope for data aggregation in the network. Data aggregation helps eliminate redundancy, and reduces the amount of data that needs to be sent to the sink or the base station.

All these features need to be taken into account when studying routing and network design wireless sensor networks.

3. Energy optimizations in routing and related problems

In this section we look at some of the important results on the problem of energy efficient routing in wireless sensor networks. Since sensor nodes are highly energy constrained, it is essential to choose the most energy efficient routes for transferring data from the source nodes to the sink nodes. With reference to the application classes discussed in 1.2, this problem is more relevant to the data gathering sensor networks than the event detection sensor networks. A seminal work in this context was presented in Chang and Tassiulas (1999); Chang and Tassiulas (Mar 2000). This work was later extended in Zussman and Segall (2003). The optimization techniques used in these works use some of the well known results on network flows from Ahuja, Magnanti and Orlin (1993). While the work in Bhardwaj and Chandrakasan (2002) is not strictly related to energy efficient routing, we discuss it in this section because it uses the same network flow optimization tools as in the above mentioned works for determining the upper bound on the lifetime of a sensor network.

In the following subsections we present an overview of these papers and some other related works on energy related optimizations in routing.

3.1 Routing for maximum system lifetime by Chang and Tassiulas

In Chang and Tassiulas (1999), the authors consider the problem of choosing routes between a set of source nodes and a set of sink nodes of an ad-hoc network so that the time until the first battery expires, is maximized. The authors note that choosing a route that results in minimum total energy expenditure as in Baker and Ephremides (1981);

Ephremides et al. (1987); Ettus (1998); Gallager Humblet and Spira (1979); Meng and Rodoplu (1998); Rodoplu and Meng (1998); Shepard (1995); Singh, Woo and Raghavendra (1998) is not always desirable because some of the nodes may have excessive relaying burden, and hence these nodes may expire too soon. This in turn could lead to loss of connectivity. To overcome this problem, the authors suggest that the routes should be chosen with the ultimate objective of maximizing the time until the first battery expires. For achieving this objective, the minimum energy paths are not necessarily the best choices.

Let E_i (in Joules) be the initial battery energy of node i, and let the node generate information at a rate of Q_i bits per second. Let S_i denote the set of nodes that can be reached by node i, and if $j \in S_i$, then let e_{ij} denote the energy required to transmit a packet from node i to node j. Let q_{ij} be the rate at which information flows from node i to node j along the link (i, j). Thus the original network topology (set N) can be thought of as a flow network with a set of source nodes (set S) and a set of sink nodes (set D) connected by a set of intermediate or relay nodes. Such flow networks have been the focus of many studies when the objective is to maximize the overall flow from the sources to the sink nodes (see Cormen, Leiserson and Rivest (2001); Ahuja, Magnanti and Orlin (1993) for more details). The flow conservation requirement at each node i that needs to be satisfied is as follows.

$$\sum_{j:i \in S_j} q_{ji} + Q_i = \sum_{k \in S_i} q_{ik}, \forall i \in N - \{D\} \tag{1.1}$$

$$\sum_{j:i \in S_i} q_{ji} = \sum_{k \in S_i} q_{ik}, \forall i \in N - \{S, D\} \tag{1.2}$$

where the first constraint says that for a node that is not a destination node, the sum of the rate at which information is received by a node and the rate at which information is generated by the node should be equal to the rate at which information is transmitted by the node. The second constraint is a special case of the first constraint when applied to nodes that are pure relays. For such nodes, the information generation rate Q_i is zero. The objective function to be maximized subject to the above constraint is the system lifetime. Or equivalently, we must determine the network flow components on all the links, i.e., $\mathbf{q} = \{q_{ij}\}$ which maximize the following, subject to (1.1) and (1.2).

$$T = T_{sys}(\mathbf{q}) = \min_{i \in N} \frac{E_i}{\sum_{j \in S_i} e_{ij} q_{ij}} \tag{1.3}$$

The above problem can be re-formulated as the following optimization problem.

$$\begin{aligned} \text{Maximize} \quad & T \\ \text{subject to:} \quad & \hat{q}_{ij} \geq 0, \qquad \forall j \in S_i, \forall i \in N - D, \\ & \sum_{j \in S_i} e_{ij}\hat{q}_{ij} \leq E_i, \qquad \forall i \in N - D, \\ & \sum_{j : i \in S_j} \hat{q}_{ji} + TQ_i = \sum_{k \in S_i} \hat{q}_{ik}, \qquad \forall i \in N - D. \end{aligned} \tag{1.4}$$

where $\hat{q}_{ij} = Tq_{ij}$ is the amount of information transfered from node i to node j along link (i, j) during time T. The above problem needs to be solved to determine $\{\hat{q}_{ij}\}$, i.e., the flow components along each link, to maximize the system lifetime (T). While this is a simple linear programming problem in $\hat{q}_{ij}$, the real challenge lies in designing a distributed algorithm to solve a lifetime maximization routing problem such as the one above. This is because using a centralized protocol for making the routing decisions is not a scalable approach, especially in large sensor networks. In addition, the control packet overheads of such an algorithm should also be low in order to make judicious use of the scarce battery resources of the nodes. An identical energy efficient lifetime maximization problem has been studied in Kalpakis, Dasgupta and Namjoshi (2002). However the authors of Kalpakis, Dasgupta and Namjoshi (2002) do not develop any distributed algorithms to solve the lifetime maximization routing problem.

In Chang and Tassiulas (1999), the authors propose two heuristic distributed algorithms to solve this problem. The first of the two algorithms is called the flow redirection algorithm. It makes use of the fact that a necessary condition for lifetime maximization is that *if the minimum lifetime over all the nodes is maximized then the minimum lifetime of a node along each path from a source to the destination has the same value as the other paths* (see Theorem 1 in Chang and Tassiulas (1999)). Each path originating from a node is associated with the smallest lifetime of the node along that path. This lifetime is computed based on $\{\hat{q}_{ij}\}$ for the current iteration. Thus the lifetime of a path is the lifetime of the shortest living node along that path, because when the shortest living node along the path expires, the path breaks down.

The intuition behind the above theorem is that if the minimum lifetime of a node along two paths is different, then we can increase the lifetime of the node with the shorter life by re-directing some of its traffic to the other path. Using this necessary condition, the authors propose a heuristic distributed algorithm that iteratively uses flow redirection along routes (by adjusting $\{\hat{q}_{ij}\}$) to maximize the minimum node lifetime. During each iteration, each node i compares routes based on the current lifetime of the shortest living node along those routes. During the next iteration node i redirects a part of the flow from a shorter living route to a longer living route by changing $\hat{q}_{ij}$ over its outgoing links. Thus the nodes attempt to "balance" the routing load over all the routes.

In the second algorithm, the authors use the heuristic of using routes with the higher residual energy. Distributed Bellman-Ford algorithm is used with the reciprocal of the residual energy as the routing metric. This algorithm performs better load balancing than the flow redirection algorithm because it takes into account the current status of the node energies by looking at the current residual energy of the nodes.

One of the limitations of this work is that since the algorithms are based on heuristics, they may not always converge to the global optimum. Another important point to note is that the work deals with pure routing and does not take into account the possibility of data aggregation at the intermediate nodes, a characteristic feature of several data gathering sensor networks.

In Chang and Tassiulas (May 2000), the authors have extended this work to obtain an approximate solution for this routing problem. The work in Chang and Tassiulas (1999) has also been extended to a multi-commodity flow problem by the same authors in Chang and Tassiulas (Mar 2000). In this case, instead of a single commodity flowing from a set of source nodes to a set of destination nodes, there is more than one commodity involved.

3.2 Energy efficient routing by Zussman and Segall

In Zussman and Segall (2003), the authors have formulated a lifetime maximization problem identical to that in Chang and Tassiulas (1999). However, the authors consider one more constraint; that of limited bandwidth. The authors study the problem of routing for maximum lifetime when the nodes have limited bandwidth in addition to limited battery energy. This is particularly true for disaster recovery ad-hoc networks consisting of smart badges. These badges are expected to have bandwidth of a few kilobits per second. Other than the bandwidth constraint, the rest of the flow conservation constraints are identical to Chang and Tassiulas (1999). The authors in Michail and Ephremides (2000) have considered a similar routing problem in the context of connection-oriented networks.

However in that work, the authors have developed heuristic algorithms instead of optimal algorithms.

The authors formulate a lifetime maximization problem identical to (1.4) along with the following capacity constraint.

$$\sum_{k\in S_i} \hat{q}_{ki} + \sum_{j\in S_i} \hat{q}_{ij} \leq T, \qquad \forall i \in N - \{S, D\}. \tag{1.5}$$

which means that the total flow through a node cannot exceed the maximum node capacity (normalized to 1).

This is a linear programming problem in $\{\hat{q}_{ij}\}$, but as in Chang and Tassiulas (1999), the challenge lies in designing a distributed algorithm to solve the problem. Unlike Chang and Tassiulas (1999) where the authors use heuristics, the authors in Zussman and Segall (2003) provide optimal algorithms along with their distributed implementations. The authors however make a simplifying assumption about the communication model. The authors assume that the nodes do not use power control when communicating with their neighboring nodes, i.e., each node i uses a fixed power level e_i when communicating with its neighbors. With this model, e_{ij} in Chang and Tassiulas (1999) is replaced by e_i[1]. With this assumption, the authors then break down the lifetime maximization problem into two loops in the algorithm. In the inner loop, the authors consider the original maximization problem in (1.4) without taking into account (1.5). For a given T, a max flow algorithm can be used to determine if there exists a feasible flow, i.e., $\{\hat{q}_{ij}\}$ which satisfies all the constraints. This is possible to do because of the assumption of $e_{ij} = e_i$. Any standard distributed implementation of a max flow algorithm (e.g., preflow-push algorithm Cormen, Leiserson and Rivest (2001)) can be used for determining the feasibility of a given T. The outer loop of the algorithm begins by checking for the feasibility of $T = T_{max}$ for the first iteration. If $T = T_{max}$ is not feasible, $T = T_{max}/2$ is checked in the next running of the outer loop. Similarly for every running of the outer loop the algorithm uses binary search to further refine the subsequent values of T. Use of binary search ensures $O(\log T_{max})$ number of iterations for determining the optimal T. Here T_{max} represents the maximum possible value of network lifetime which is upper bounded by n times the lifetime of a single battery, where n is the total number of nodes in the network.

In the above problem formulation, the authors associate a fixed amount of energy e_i with node i for packet transmission. Since other than the

---

[1] If this assumption is made for the problem in Chang and Tassiulas (1999), the necessary condition in Theorem 1 also becomes a sufficient condition, and then the algorithms proposed in Chang and Tassiulas (1999) converge to the global optimum.

source nodes all the other nodes act as relay nodes, instead of accounting for transmission energy and reception energy separately for each packet, the authors absorb the energy spent on receiving a packet in the transmission energy e_i. Thus e_i actually represents the energy spent on relaying a packet.

However, one of the most important limitations of transceivers used in sensor nodes is their idle mode energy consumption. Transceivers spend a considerable amount of energy when their radio is in idle mode, i.e., neither transmitting nor receiving, and sometimes this energy is as high as the energy spent on transmission or reception (see Shih et al. (2001)). As a result, when the transmissions and receptions are not perfectly synchronized, the nodes continue to spend energy on idle listening. This is especially true for a multi-hop network where a relay node does not know beforehand when it is going to receive the next packet. The simplistic model of associating a fixed amount of energy with each packet transmission without accounting for idle mode energy is an idealistic scenario. While it is true that taking into account the underlying MAC protocol makes the analysis difficult to handle, the fact that the MAC layer has a large impact on the energy expenditure of a sensor node cannot be ignored. A natural extension of Zussman and Segall (2003) would be to formulate an identical lifetime maximization problem by accounting for idle mode energy expenditure.

3.3 Bounding the lifetime of a sensor network by Bhardwaj and Chandrakasan, a related problem

The authors in Bhardwaj and Chandrakasan (2002) study the problem of obtaining bounds on the lifetime of a sensor network. The authors use similar network flow tools as in Zussman and Segall (2003). However they also take into account the possibility of data aggregation at some of the nodes. With data aggregation, the flow conservation constraints have to be modified at the aggregating nodes. However most of the other constraints are identical to (1.4). By formulating a lifetime maximization problem as in (1.4), we obtain a linear programming problem that can be solved for a given network topology. The solution of this problem, i.e., the optimum $\{\hat{q}_{ij}\}$ provides an upper bound on the lifetime of the network. However, as the authors themselves state, it is difficult to design a distributed routing protocol that achieves these bounds.

As in 1.3.2, this work assumes perfect transmitter-receiver synchronization in energy analysis. This work can be extended in two directions; a distributed routing protocol that achieves flow rates corresponding to the optimum solution of the problem, $\{\hat{q}_{ij}\}$ can be developed, and the problem formulation can be modified to take into account the underlying MAC.

3.4 Other related work

The problem of network lifetime maximization has been addressed in several other works which are not related to routing, but which use network flow tools. In Srinivasan et al. (2002), the authors formulate an optimization problem by associating a utility function with every source node. The utility function is an increasing and concave function of the flow rate out of the source node. The objective is to determine $\{\hat{q}_{ij}\}$ that maximizes the sum of the utilities of all the sources while ensuring a certain minimum lifetime. There is also an upper bound on the allowable source rates. The authors use a penalty function based approach for the system utility maximization, and also propose a distributed algorithm called Optimal Rate Splitting and Allocation (ORSA) that can be implemented at the source nodes to determine the optimum source rates.

In Shah and Rabaey (2002), the authors address the problem of lifetime maximization by picking the next hop nodes in a probabilistic fashion. This is a heuristic algorithm, and the probability of choosing a neighbor node as the next hop node is proportional to the inverse of the cost of the link to that node. The cost of a link equals the energy spent on transmitting a packet on that link. This form of randomness in choosing the next hop node ensures some level of load balancing which is better than always choosing the minimum energy route, because the latter results in quick depletion of energy resources along the minimum energy route.

4. Cost optimizations in network design

Wireless sensor networks are characterized by their high node density and possibility of data aggregation. Since node measurements from neighboring nodes are expected to be correlated, it is possible to perform in-network aggregation of the measured data so as to reduce the amount of data that needs to be sent to the base station. Besides, the high node density requires hierarchical management of the network for better scalability of protocols. Organizing nodes into clusters is one of the ways to achieve these objectives. Several alternatives are available when designing clustered sensor networks. For example, the network could consist of multiple types of nodes, such that the cluster head responsibilities can be assigned to one type of sophisticated nodes, while the rest of the simpler nodes perform sensing. Within each cluster the nodes could use single hop or multi-hop communication to reach the cluster head nodes. The radius of communication for multi-hopping is another parameter at the designer's disposal. These and other structural characteristics of sensor networks need to be taken into account when designing the sensor networks. In this section, we survey some of the work done on energy and cost efficient design of wireless sensor networks.

With reference to the application classes discussed in 1.2, this kind of networks fall under the category of data gathering networks. For such networks, it is possible to model the data gathering process as a set of discrete cycles. During each cycle, the nodes send their measured data to the cluster head nodes which perform some data aggregation, and then send the aggregated data to the base station.

4.1 Design optimizations in homogeneous sensor networks

We first begin by looking at the design of homogeneous sensor networks. In a homogeneous sensor network, all the nodes are identical in terms of their hardware and battery energy.

4.1.1 A single hop homogeneous clustered network, LEACH by Heinzelman et al. In Heinzelman, Chandrakasan and Balakrishnan (2002), a distributed data gathering protocol called LEACH, i.e., Low Energy Adaptive Clustering Hierarchy is proposed for a sensor network is in which a fixed number of homogeneous nodes are distributed randomly over a region. There is a remote base station that is located outside the region. Nodes are organized into clusters, and the cluster head nodes are chosen from among the sensor nodes. During each data gathering phase, the nodes send their measured data to the closest cluster head node through a direct transmission. The cluster head node aggregates the received packets into a single packet, and transmits it to the remote base station. Since the cluster head nodes carry the burden of long range transmissions to the base station, they are likely to drain their battery before other nodes. Hence in order to ensure some form of load balancing, the role of the cluster head nodes is rotated randomly and periodically over all the nodes in the network. Since the nodes are homogeneous, all the nodes have the hardware required for performing long range transmissions to the remote base station, and for performing data aggregation computations.

The question that the authors address under these settings is: what is the optimum number of cluster head nodes required that minimizes the average energy expenditure of each node during a single data gathering cycle? For this, the authors obtain an expression for the energy spent in the entire network during each data gathering cycle, and then minimize it with respect to the number of cluster heads. Because of cluster head rotation, there is a more or less uniform drainage of energy over the entire network, and hence the authors seek to minimize the network wide energy expenditure. Note that this effectively means minimizing the required battery energy of each node for a given system lifetime. The larger the number of cluster heads, the smaller the distance over which the nodes have to transmit to reach the cluster head nodes; however the higher the number of long range energy intensive transmissions to the remote base station. Hence there is an inherent trade-off which means that there is an optimum number of cluster head nodes. The optimum number of cluster heads is obtained by differentiating the average network-wide energy expenditure with respect to the number of cluster heads, and equating the resulting expression to zero.

One of the most important characteristics of LEACH is node homogeneity. In order to use cluster head rotation, it is necessary that every node be equipped with complex hardware for long range communication with the remote base station. This results in an increased hardware cost of the overall network. Thus while the authors minimize the battery energy requirements of each node, the hardware cost requirements are not taken into account in the problem analysis. The data aggregation model that is used by the authors also leaves much to be desired. In general, for most applications, it is not reasonable to assume that irrespective of the size of a cluster (which is a variable over which optimization is performed), the data packets of all the nodes in that cluster can be aggregated into a single packet of fixed size. More elaborate data aggregation models which take into account the extent of correlation in the measured data as discussed in Mhatre and Rosenberg (2003) should be considered.

In a related paper by Lindsey and Raghavendra (2001), the authors propose a data gathering scheme called PEGASIS, i.e., Power-Efficient Gathering in Sensor Information Systems for a homogeneous sensor network. In this scheme there is a single cluster head node, and this role of cluster head is rotated periodically over all the nodes as in LEACH. The difference between PEGASIS and LEACH is that the authors of PEGASIS assume an aggregation model in which nodes are allowed to aggregate data along each hop. Thus each node receives packets from the nodes which are farther from the cluster head, and aggregates these packets along with its own packet to produce a single packet which is then sent to the next hop node. The aggregation model used in PEGA-

SIS is even more restrictive than the model used in LEACH. PEGASIS also requires proper scheduling of transmissions among all the nodes so that hop-by-hop aggregation is possible.

4.1.2 Minimizing communication costs by Bandyopadhyay and Coyle. In Bandyopadhyay and Coyle (2003), the authors consider a sensor network in which the nodes are distributed over a circular region, and the base station is located at the center of the region. The nodes are homogeneous, and are organized in clusters. Each node has the same communication radius. The nodes use multi-hop communication to reach the cluster head node. The authors assume a data gathering network model in which the nodes send their measured data to their respective cluster heads during each data gathering cycle. The cluster head node aggregates the received packets into a single packet, and then sends the aggregated packet to the central base station using multi-hopping. While in LEACH the communication paradigm is single hopping for communication between the nodes and their cluster heads, and between the cluster heads and the base station, in this case, the communication paradigm is multi-hopping. The authors use tools from stochastic geometry to obtain an expression for the energy spent in the entire network during each data gathering cycle, and minimize this to obtain the optimum number of cluster head nodes. No cluster head rotation is used. The energy minimization problem is identical to the LEACH energy minimization problem in the sense that the authors minimize the network-wide energy expenditure with respect to the number of cluster heads.

An important observation that can be made about the scheme is that since the nodes and the cluster heads use multi-hopping, the nodes around the cluster heads and the nodes around the base station have the highest energy drainage burden due to excessive relaying of packets. While in LEACH role rotation ensures uniform energy drainage over all the nodes, this scheme suffers from the problem of hot spot formation around the cluster head nodes and the central base station. As a result, it is the energy expenditure of the nodes in these hot-spots that determines the lifetime of the system, and this observation needs to be taken into account in the minimization problem. The work also suffers from a restrictive data aggregation model like the work in 1.4.1.1.

4.2 Design optimizations in heterogeneous sensor networks

In the previous section, we looked at some of the design optimizations in homogeneous sensor networks. In such networks, the primary objective is minimizing the battery expenditure of each node for a given lifetime. However, there is another class of sensor networks which uses two or more types of nodes. For example, with two types of nodes the type 0 nodes act as sensor nodes, while type 1 nodes act as cluster head nodes. Most of the complex hardware and software functionality can be embedded in a few type 1 nodes, while the type 0 nodes can be designed to be simple. In this section, we look at the design of such heterogeneous networks.

4.2.1 A minimum cost heterogeneous network by Mhatre et al. In Mhatre et al. (2003), the authors consider a heterogeneous clustered sensor network and a periodic data gathering network model. The base station is located outside the region of interest. There are two types of nodes; type 0 nodes which are pure sensor nodes, and type 1 nodes which are cluster head nodes. The sensor nodes use short range multi-hop communication to reach the closest cluster head node. The cluster head nodes receive packets from all the nodes in their respective clusters, aggregate the received packets into a single packet, and transmit the aggregated packet to the remote base station using a direct transmission. Since the cluster head nodes require the hardware to communicate over larger distances as compared to the sensor nodes, the hardware cost of a cluster head node, α_1 is larger than the hardware cost of the sensor node, α_0. Since the type 0 nodes use multi-hopping to reach the closest type 1 node, hot spots are formed around the type 1 nodes. The type 0 nodes which are within these hot spots, i.e., within one hop of the type 1 nodes, are called *critical nodes.* The critical type 0 nodes expire before other type 0 nodes because all the packets in their cluster have to be relayed by them over the last hop, and this results in a higher relaying burden.

In order to ensure a lifetime of at least T data gathering cycles, it is necessary that the type 1 nodes and the critical type 0 nodes have sufficient battery energy to last for T cycles. The authors obtain expressions for the energy expenditure of both types of nodes, and then determine the corresponding battery requirements, E_i. They then formulate an optimization problem with the following cost function:

$$C(\bar{\lambda}, \bar{E}) = \lambda_0(\alpha_0 + \beta E_0) + \lambda_1(\alpha_1 + \beta E_1) \tag{1.6}$$

In the above cost function, λ_i is the intensity (number of nodes per unit area) of type i nodes, and β is a proportionality constant so that βE_i is the cost of the battery of type i node. Thus we note that unlike the homogeneous network where the objective function to be minimized is simply the battery energy, in the case of a heterogeneous network the objective function to be minimized involves battery energy as well as the hardware cost of the multiple types of nodes.

As in Bandyopadhyay and Coyle (2003), the authors use tools from stochastic geometry to obtain expressions for E_i. The cluster head nodes spend energy on receiving packets from all the nodes in their respective clusters, aggregating the packets, and then making a long range transmission to the base station. The energy expenditure of a critical type 0 node is obtained by first determining the average relaying load on a critical node. The relaying load on a critical node is simply the ratio of the average number of nodes in the cluster minus the average number of critical nodes, to the average number of critical nodes. This ratio is the average number of packets that a critical node must relay. Once the relaying load on a critical node is determined, the required battery energy of type 0 nodes, E_0 is known. The minimum lifetime requirement of T data gathering cycles results in the following inequality constraint.

$$\frac{E_1}{P_1} = \frac{E_0}{P_0} \geq T \tag{1.7}$$

where P_0 is the average energy expenditure of a critical type 0 node, and P_1 is the average energy expenditure of a type 1 node during a single data gathering cycle. The authors also take into account the connectivity-coverage requirements in the form of an additional constraint which requires that the total node intensity $\lambda_0 + \lambda_1$ be greater than a threshold to ensure node connectivity and area coverage with a probability of at least $1 - \epsilon$. Node connectivity is required for multi-hop communication to be possible. Minimizing the cost function in (1.6) with respect to λ_0 and λ_1 yields the optimum cluster head and sensor node intensities. The authors use Karush-Kuhn-Tucker theorem to minimize the cost function under the equality and inequality constraints.

An important limitation of the results obtained in Mhatre et al. (2003) is that the authors assume an ideal MAC in analyzing the problem, i.e., they assume that there is no energy wasted by the nodes on idle listening, and that there are no packet collisions. While this is reasonable in the case of single hop clusters as in LEACH, it is difficult to ensure in the case of multi-hop clusters. This is because the transmissions and receptions of all the nodes over all the hops need to be synchronized in a multi-hop cluster.

4.2.2 Optimum mode of communication in a heterogeneous network, Mhatre and Rosenberg. It is well-known that in general, multi-hop communication is preferable to single hop communication, since the signal strength over distance d falls as $1/d^k$, $k \geq 2$. However in practical transceivers, each packet transmission is also associated with constant overheads due to the energy spent in the digital circuitry. In Mhatre and Rosenberg (2003), the authors consider a heterogeneous network as in Mhatre et al. (2003). However, instead of assuming that the nodes communicate with a fixed radius of communication, the authors let the radius of communication be another variable in the optimization problem. The optimization problem is formulated along the same lines as (1.6) with a minor modification that in addition to the node intensities, the radius of communication is also a variable. There are two constraints on the communication radius. Firstly, it should be greater than or equal to the minimum radius required for node connectivity in order that multi-hop communication be possible. Secondly, the communication radius should be smaller than the average radius of each cluster. The second constraint is required because when the communication radius becomes equal to the average radius of a cluster, the nodes can communicate with the cluster head using a single hop transmission, and such a single hop clustered network can be analyzed separately as is done in Mhatre and Rosenberg (2003).

During each data gathering cycle, nodes send their measured data to their respective cluster head nodes which aggregate the received packets into a single packet, and transmit it directly to the remote base station. The Karush-Kuhn-Tucker theorem is used for cost minimization as in Mhatre et al. (2003). In the solution of the optimization problem, if it turns out that the optimum radius of communication is equal to the average radius of a cluster, then clearly single hopping is the optimum choice for in-cluster communication. If not, then the optimum mode of communication is multi-hopping with a radius of communication given by $\hat{R}$ as follows.

$$\hat{R} = \left(\frac{4l}{\mu(k-2)} \right)^{1/k}, \tag{1.8}$$

where we assume a radio model in which the energy required to transmit a packet over distance d is $l + \mu d^k$. Here k is the propagation loss exponent, and l is the fixed amount of energy that is spent in the digital circuitry during the packet transmission. Thus, for heterogeneous networks, there is an optimum radius of communication $\hat{R}$ given by (1.8) which depends only on the radio parameters of the transceiver and the propagation loss exponent.

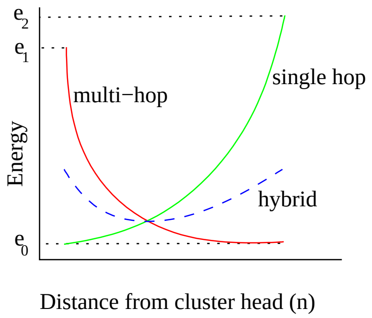

Figure 1.2. Hybrid Communication Mode.

The authors also propose a hybrid mode of communication in which the nodes periodically alternate between single hopping and multi-hopping for in-cluster communication. The intuition behind this idea is that when nodes use single hopping to reach the cluster head node, the nodes that are farthest from the cluster head node have the highest energy burden. On the other hand, when the nodes use multi-hop mode, the nodes that are closest to the cluster head node (within one hop) have the highest energy burden due to excessive packet relaying. Hence a periodic mode rotation between single hopping and multi-hopping leads to a more uniform energy drainage pattern. The exact fraction of time for which each of the modes is to be sustained is determined so that the energy expenditure profile of the hybrid mode has the same value at both its end points (see Figure 1.2). This ensures that the nodes that are burdened by single hopping and the nodes that are burdened by multi-hopping expire at about the same time. The cost minimization problem is again solved using the Karush-Kuhn-Tucker theorem.

4.3 Homogeneous versus heterogeneous networks by Mhatre and Rosenberg

In Heinzelman, Chandrakasan and Balakrishnan (2002); Bandyopadhyay and Coyle (2003); Mhatre et al. (2003); Mhatre and Rosenberg (2003), the authors begin by studying either a homogeneous or a heterogeneous sensor network, and then optimizing the corresponding network cost (which is just the battery cost for homogeneous networks and the battery plus the hardware cost for heterogeneous networks). However they do not provide any guidelines as to which is the best of the two networks; homogeneous or heterogeneous. Mhatre and Rosenberg address this problem in Mhatre and Rosenberg (2004) where they compare homogeneous and heterogeneous networks based on the overall cost of the network.

The authors use the cost metric given by (1.6) for the purpose of comparison. With a homogeneous network, for example LEACH, the uniform energy drainage due to role rotation ensures that the required battery energy in each node is minimized. However it also requires each node to have complex hardware to act as a cluster head. Thus in the case of LEACH, the overall cost of the network, $f_1(\alpha_0, \alpha_1, \beta)$ is as follows.

$$f_1(\alpha_0, \alpha_1, \beta) = n_0(\alpha_1 + \beta E) \tag{1.9}$$

where n_0 is the number of nodes in the network, and α_1 is the hardware cost of a cluster head node. Due to role rotation, each node has to be capable of transmitting directly to the remote base station, and perform other duties of a cluster head, and therefore the hardware cost of each node is α_1. On the other hand, the cost of the corresponding heterogeneous network, $f_2(\alpha_0, \alpha_1, \beta)$ is as follows.

$$f_2(\alpha_0, \alpha_1, \beta) = n_0(\alpha_0 + \beta E_0) + n_1(\alpha_1 + \beta E_1) \tag{1.10}$$

Note that in the above equation the complex hardware functionalities are embedded in only a few nodes (n_1 cluster head nodes), and therefore the overall hardware cost of the system is low. However since there is no role rotation, the non-uniform energy drainage results in a higher battery energy in each node. Thus there is a trade-off between homogeneous and heterogeneous networks in terms of the cost of the battery and the hardware.

In Mhatre and Rosenberg (2004), the authors first determine the minimized costs of both homogeneous and heterogeneous networks for given settings. Then they determine the difference between these minimized costs, i.e., (1.9) $-$ (1.10), and this serves as a guideline for the designers to choose between a homogeneous and a heterogeneous network. The authors also propose a multi-hop generalization of LEACH called M-LEACH. They note that in the original LEACH scheme, the nodes use single hopping to reach their respective cluster head nodes. However the nodes could use multi-hopping to reach the cluster head nodes to save on battery energy by avoiding distant transmissions to the cluster heads. M-LEACH is still a scheme for homogeneous networks, but it allows for multi-hopping within the cluster.

4.4 Other related work

In Chiasserini et al. (2002), the authors consider a single hop clustered sensor network in which the lifetime of the network is defined as the time until the first cluster head node expires. The number of sensor nodes and the number of cluster heads is fixed, and is given. The authors address the problem of optimal assignment of nodes to the cluster heads so as to maximize the lifetime of the network. There is no role rotation, and the topology is static. Nodes are assigned to the cluster heads so as to balance the load on all the cluster head nodes. The mode of communication within the cluster heads is single hopping. It is assumed that the location of all the nodes is known, and this information is used to determine the node assignment policy for each cluster head node. However, the authors have not developed a distributed protocol for solving this problem.

5. Conclusions

In this survey paper, we provided an overview of some of the recent work on energy and cost optimizations in wireless sensor networks. Sensor nodes are highly energy constrained, and energy efficiency is of prime importance at all the layers of protocol stack. Different network design issues surface depending on the kind of application involved. In this survey, we restricted ourselves mainly to those applications which are of data gathering type. We focused our attention on two important aspects of sensor networks, namely routing and design optimizations. In the context of routing optimizations, we looked at some of the important papers on energy efficient routing for maximizing the system lifetime. Several tools from the theory of network flows were used to tackle these optimization problems. We then looked at some of the important works on design related optimization problems in sensor networks. We focused our attention on clustered sensor networks, and the problem of cost (battery plus hardware) minimization. We noted that several optimization tools and techniques are useful in designing and dimensioning of wireless sensor networks.

Acknowledgments

This work was supported in part by an NSF grant (contract no. 0087266).

References

Ahuja, R.K., Magnanti, T.L. and Orlin, J.B. *Network Flows*, Prentice Hall, 1993.

Akyildiz, I.F., Su, W., Sankarsubramaniam, Y. and Cayirci, E. Wireless Sensor Networks: a survey. *Computer Networks*, Vol. 38, pp. 393-422, March 2002.

Baker, D.J. and Ephremides, A. The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm. *IEEE Transactions on Communications*, vol. COM-29, no. 11, pp. 56-73, January 1981.

Bandyopadhyay, S. and Coyle, E. An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. *Proceedings of IEEE INFOCOM'03*, San Francisco, CA, April 2003.

Bhardwaj, M. and Chandrakasan, A. Bounding the lifetime of sensor networks via optimal role assignments. *IEEE INFOCOM'02*, New York, NY, Jun 2002.

Chang, J. and Tassiulas, L. Routing for maximum lifetime in wireless ad-hoc networks. *37th Annual Allerton Conference on Communication, Control and Computation*, Sep. 1999.

Chang, J. and Tassiulas, L. Energy conserving routing in wireless ad-hoc networks. *IEEE INFOCOM'00* Tel Aviv, Israel, Mar 2000.

Chang, J. and Tassiulas, L. Fast approximate algorithms for maximum lifetime routing in wireless ad-hoc networks. *IFIP-TC6 Networking 2000, LNCS*, Vol. 1815, Springer, May 2000.

Chiasserini, C.F., Chlamtac, I., Monti, P. and Nucci, A. Energy Efficient Design of Wireless Ad Hoc Networks. *Proceedings of IFIP-TC6 Networking 2002, LNCS*, Vol. 2345, Springer, May 2002.

Cormen, T.H., Leiserson, C.E. and Rivest, R.L. *Introduction to Algorithms*, 2001, Prentice Hall.

Ephremides, A., Wieselthier E.J. and Baker, D.J. A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling. *Proceedings of IEEE*, vol. 75, no. 1, pp. 56-73, January 1987.

Ettus M. System Capacity, Latency and Power Consumption in Multihop-routed SS-CDMA Wireless Networks. *Proceedings of IEEE Radio and Wireless Conference (RAWCON) 98*, Colorado Springs, CO, Aug 1998, pp. 55-58.

Gallager, R.G., Humblet, P.A. and Spira, P.M. A Distributed Algorithm for Minimum Weight Spanning Trees. Technical Report LIDS-P-906-A, Laboratory of Information Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, October 1979.

Heinzelman, W., Chandrakasan, A. and Balakrishnan, H. An Application-Specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, October 2002.

Kalpakis, K., Dasgupta, K. and Namjoshi, P. Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks. *Proceedings of IEEE Networks'02 Conference*, 2002.

Lindsey, S. and Raghavendra, C.S. PEGASIS: Power-Efficient Gathering in Sensor Information Systems. *IEEE Aerospace Conference* (Big Sky, MT, Mar. 9-16, 2002), Vol. 3, p. 3-1130.

Meng, T.H. and Rodoplu, V. Distributed Network Protocols for Wireless Communication. *Proceedings of 1998 IEEE Symposium on Circuits and Systems*. ISCAS'98, Monterey, CA, June 1998, vol. 4, pp. 600-603.

Mhatre, V. and Rosenberg, C. Design guidelines for wireless sensor networks: Communication, clustering and aggregation. *Ad Hoc Networks Journal, Elsevier Science*, Volume 2, Issue 1, pp 45-63, 2003.

Mhatre, V. and Rosenberg, C. Homogeneous vs heterogeneous sensor networks: A comparative study. In the proceedings of *International Conference on Communications (ICC 2004)*, Paris, France, June 2004.

Mhatre, V., Rosenberg, C.P., Kofman, D., Mazumdar, R.R. and Shroff, N.B. A minimum cost surveillance sensor network with a lifetime constraint. To appear in *IEEE Transactions on Mobile Computing*. Available at `http://web.ics.purdue.edu/~mhatre/lifetime.pdf`.

Michail, A. and Ephremides, A. Energy Efficient Routing for Connection-Oriented Traffic in Ad-Hoc Wireless Networks. *Proceedings of IEEE PIMRC'00*, Sep 2000.

Rodoplu, V. and Meng, T.H. Minimum Energy Mobile Wireless Networks. *Proceedings of 1998 IEEE International Conference on Communications, ICC'98*, Atlanta, GA, June 1998, vol. 3, pp. 1633-1639.

Shah, R.C. and Rabaey, J.M. Energy Aware routing for Low Energy Ad Hoc Sensor Networks. *Proceedings of IEEE WCNC'02*, Mar. 2002.

Shepard, T. Decentralized Channel Management in Scalable Multihop Spread Spectrum Packet Radio Networks. Technical Report MIT/LCS/TR-670, Massachusetts Institute of Technology Laboratory for Computer Science, July 1995.

Shih, E., Cho, S., Ickes, N., Min, R., Sinha, A., Wang A. and Chandrakasan, A. Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. *Proceedings of ACM MobiCom'01*, Rome, Italy, July 2001, pp. 272-286.

Singh, S., Woo, M. and Raghavendra, C.S. Power-aware Routing in Mobile Ad Hoc Networks. *Proceedings of Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking*, Dallas, TX, Oct. 1998, pp. 181-190.

Srinivasan, V., Chiasserini, C.F., Nuggehalli, P. and Rao, R. Rate allocation and traffic splits for energy efficient routing in ad hoc networks. *IEEE INFOCOM'02*, New York, NY, Jun 2002.

Tilak, S., Abu-Ghazaleh, N.B. and Heinzelman, W. A Taxonomy of Wireless Sensor Network Communication Models. *Mobile Computing and Communication Review*, April 2002, Vol. 6, No. 2.

Zussman, G. and Segall, A. Energy efficient routing in ad hoc disaster recovery networks. *IEEE INFOCOM'03*, San Francisco, CA, Apr 2003.