Power Balanced Coverage-Time Optimization for Clustered Wireless Sensor Networks

Tao Shu
Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721, USA
tshu@ece.arizona.edu

Marwan Krunz
Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721, USA
krunz@ece.arizona.edu

Sarma Vrudhula
Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721, USA
vrudhula@ece.arizona.edu

ABSTRACT

We consider a wireless sensor network in which sensors are grouped into clusters, each with its own cluster head (CH). Each CH collects data from sensors in its cluster and relays them to a sink node directly or through other CHs. The *coverage time* of the network is defined as the time until one of the CHs runs out of battery, resulting in an incomplete coverage of the sensing region. We study the maximization of coverage time by balancing the power consumption of different CHs. Using a Rayleigh fading channel model for inter-cluster communications, we provide optimal power allocation strategies that guarantee (in a probabilistic sense) an upper bound on the end-to-end (inter-CH) path reliability. Our allocation strategies account for the interaction between routing and clustering by considering the impacts of intra- and inter-cluster traffic at each CH. Two mechanisms are proposed for achieving balanced power consumption: the routing-aware optimal cluster planning and the clustering-aware optimal random relay. For both mechanisms, the problem is formulated as a signomial optimization, which can be efficiently solved using generalized geometric programming. Numerical examples and simulations are used to validate our analysis and study the performance of the proposed schemes.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless Communication*

General Terms

Algorithms, Performance, Design.

Keywords

Generalized geometric programming, signomial optimization, sensor networks, clustering, topology control, coverage time.

MobiHoc'05, May 25–27, 2005, Chicago, Illinois, USA.

1. INTRODUCTION

1.1 Motivation

Recent technological advances in micro-electronic-mechanics (MEMs) and low-power wireless communications have paved the way for the deployment of dense wireless sensor networks (WSNs). Such networks are expected to play an important role in a wide range of applications, including motion detection, environment monitoring, military surveillance and reconnaissance, etc. [1]. Mass production of low-cost sensors necessitates powering them with limited-energy, often non-rechargeable batteries [2]. This makes energy consumption a critical factor in the design of a WSN and calls for energy-efficient communication protocols that maximize the lifetime of the network subject to a given energy budget.

Sensors in a WSN are often organized into clusters, each with its own CH [2]. Instead of communicating directly with the data processing center, referred to as the *sink*, a sensor transmits data to its CH, which in turn forwards the data to the sink, either directly or via a multi-hop path through other (intermediate) CHs. This significantly reduces the battery drainage of individual sensors, which only need to communicate with their respective CHs over relatively short distances. Besides its energy-saving advantages, clustering has other desirable features related to network management, security, scalability, etc.

Expectedly, the clustering paradigm increases the burden on the CHs, forcing them to deplete their batteries much faster than in a non-clustered network. The additional energy consumption is attributed to the aggregation of intra-cluster traffic into a single stream that is transmitted by the CH and to the relaying of inter-cluster traffic from other CHs. Such relaying is sometimes desirable because of its power-consumption advantage over direct (CH-to-sink) communication. Given the high density of sensors in common deployment scenarios (a density of 20 sensors per cubic meter is not unusual [1]), the traffic volume coming from a CH can be orders of magnitude greater than the traffic volume of an individual sensor. Even though the CH may be equipped with a more durable battery than the individual sensors it serves, the large difference in power consumption between the two can lead to shorter lifetime for the CH. Once the CH dies, no communications can take place between the sensors in that cluster and the rest of the network.

For clusters of comparable coverage and node density, the intra-cluster traffic volume is roughly the same for all clusters. On the other hand, the traffic volume relayed by different CHs is highly skewed; the closer a CH is to the sink, the more traffic it has to relay. Such a "traffic implosion" situation is essentially caused by the many-to-one communication paradigm in WSNs, i.e., traffic from all sensors is eventually destined to the sink. Unless some action is taken to correct this imbalance, different CHs will drain their batteries at different times, resulting in early (partial) loss of coverage and potential partitioning of the underlying topology. Our goal in this paper is to design optimal power allocation strategies that address this imbalance by maximizing the *coverage time*, defined as the time until one CH runs out of battery[1]. These strategies deliberately offset the impact of the skewed load by appropriately adjusting the transmission distance (power) of different CHs. Because the volume of relayed traffic is also affected by the underlying routing scheme, a joint routing/clustering design methodology is needed to achieve power balance among CHs.

1.2 Related Work

Extensive research has been dedicated to the study of clustering algorithms for ad hoc and wireless sensor networks. Early clustering algorithms mainly focused on the connectivity problem [3]-[9], aiming at generating the minimum number of clusters that ensures network connectivity. In these algorithms, the election of the CH is done based on node identity [3, 4, 5], connectivity degree [6], or connected dominating set [7]-[9].

Recently, there has been increased interest in studying energy-efficient clustering algorithms, in the context of both ad hoc and sensor networks [10]-[17]. In [10], the authors proposed the LEACH algorithm, in which the CH role is dynamically rotated among all sensors in the cluster. Energy is evenly drained from various sensors, leading to improved network lifetime. A similar CH-scheduling scheme was proposed in [17] for a time-slotted WSN. In this scheme, several disjoint dominating sets are found and are activated successively. Nodes that are not in the currently active dominating set are put to sleep. A distributed algorithm was proposed to obtain a set-schedule sequence for which the network lifetime is within a logarithmic factor of the maximum achievable lifetime (obtained under an optimal sequencing of the dominating sets). In general, such rotation-based algorithms require excessive processing and communication overheads for CH re-election and broadcasting of the new CH information.

"Load-balanced" algorithms (e.g., [11]-[13]) focus mainly on balancing the intra-cluster traffic load, and ignore inter-cluster traffic. In [11], sensors are clustered according to "load-balancing" metrics, whereby the traffic volumes originating from various clusters are equalized. The authors in [12] extended the work in [11] by integrating the concept of load balancing into traditional node-id/connectivity-degree based clustering to produce a longer CH lifespan. In [13], the max-min d-cluster algorithm was proposed to extend the traditional 1-hop cluster to a d-hop cluster while generating load-balanced clusters. This extension achieves better load balancing with fewer clusters.

Distributed algorithms for organizing sensors into a hierarchy of clusters were proposed in [15, 16], with the objective of minimizing the energy spent in communicating information to the sink. It should be noted that minimizing the total energy consumption is not equivalent to maximizing coverage time, as the former criterion does not guarantee balanced power consumption at various CHs. By shifting the load from over-power-drained CHs to under-power-drained CHs, coverage time can be maximized even though the total power consumption is not necessarily minimal.

In [14], the authors proposed clustering algorithms that maximize network lifetime by determining the optimal cluster size and optimal assignment of nodes to preselected CHs. Their exhaustive-search approach assumes full knowledge of the network topology (i.e., the location of each sensor node and each CH in the network). Also, it ignores inter-cluster traffic.

The scheme in [18] incorporates the impact of inter-cluster traffic in determining the optimal location of the sink node that maximizes the topological lifetime of the network. Power-balance among CHs was not considered. To the best of our knowledge, there is no existing literature that adequately addresses power balance among CHs and provides optimal power allocation strategies that maximize the coverage time.

1.3 Main Contributions and Paper Organization

The main contributions of this paper are as follows. First, in contrast to previous "load-balanced" algorithms, we provide "power-balanced" alternatives that aim at *directly* optimizing coverage time by accounting for the interaction between clustering and routing, i.e., simultaneously taking into consideration the impacts of both intra- and inter-cluster traffic. Second, in contrast to previous algorithms, which are based on heuristics, ours is based on an analytical approach in which coverage-time maximization is formulated as a signomial optimization problem that can be efficiently solved using *generalized geometric programming* (GGP) [22, 23]. Our analysis guarantees an upper bound on the path reliability for communications between the originating CH and the sink node. Two schemes are proposed for achieving power-balanced communications: *routing-aware optimal cluster planning* and *clustering-aware optimal random relay*. The first scheme is essentially a clustering approach that is developed in the context of shortest-hop-count inter-CH routing. For this scheme, the optimal cluster size and location are obtained. The second scheme is essentially a routing strategy for "load-balanced" clustered topologies (i.e., all clusters are of the same size). According to this approach, a CH probabilistically chooses to either relay the traffic to the next-hop CH or to deliver it directly to the sink.

Numerical examples and simulations are used to validate our analysis and compare our proposed schemes with pure "load-balancing" algorithms. Our results indicate that by accounting for the interaction between clustering and routing, the proposed schemes achieve a significant reduction in energy consumption and improved coverage time.

The rest of this paper is organized as follows. In Section 2 we describe the system model and the assumptions made in the analysis. The optimization problems are formulated in Section III. In Section IV we validate our analysis using numerical examples and computer simulations. Section V

[1] Other definitions for coverage time may also be used, such as the time until $x\%$ of coverage is lost or the time until the network is partitioned. Such definitions will be considered in a future work.

concludes the paper.

2. SYSTEM MODEL AND ASSUMPTIONS

2.1 Network Model

We consider a circular sensing area **A** of radius R. The sink is located at the center, as shown in Figure 1. The circular geometry, albeit too idealistic, serves as a basis for understanding the intrinsic tradeoffs involved in a joint clustering/routing optimization framework. It has been widely used in the analysis of sensor networks; most recently in [19, 20]. The sensors are uniformly distributed across **A** with density ρ. Data captured and generated by all sensors need to be delivered to the sink. Due to energy considerations, only those sensors within the area $\{(x,y)\,|x^2+y^2 \le r_0^2\}$, where $r_0 < R$, can communicate directly with the sink; all other sensors are organized into clusters and they communicate their data through their respective CHs. Without loss of generality, we assume that each CH is located at the center of its cluster.

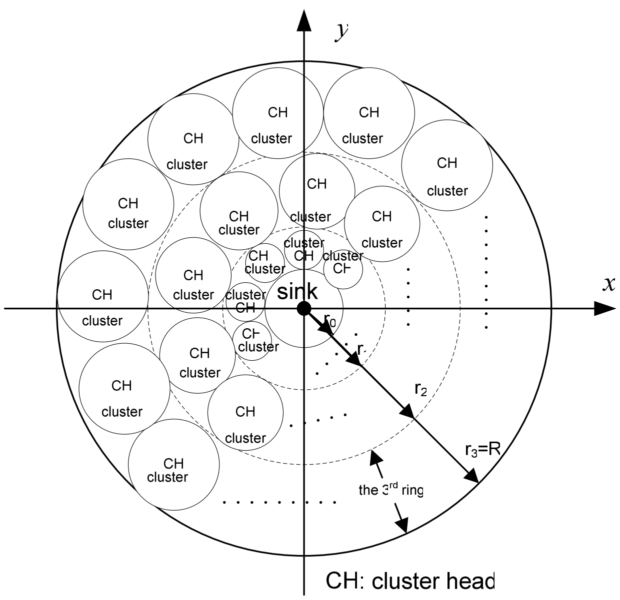

Figure 1: Network topology with three rings ($K = 3$).

The procedure for cluster formation consists of two steps: the deployment of CHs and the assignment of sensors to CHs. Because of the symmetric nature of the area **A** and the uniform distribution of sensors, the formation of clusters is also symmetric, i.e., any two clusters with the same distance from their centers to the sink should have the same coverage. Such clusters are said to be of the same type. Suppose there are K types of clusters in the network. We consider the following clustering approach: sensors whose distances to the sink fall in $(r_{i-1}, r_i]$ are organized into clusters of the ith type, where $1 \le i \le K$ and $r_0 < r_1 < \ldots < r_K = R$. As a result, the clusters of the ith type cover the ith *ring*, defined by the area $\{(x,y)\,|r_{i-1}^2 < x^2+y^2 \le r_i^2\}$. This clustering approach can be easily realized in practice, e.g., by using pilot signals that are broadcasted by the sink. Accordingly, the CHs of the ith ring are placed evenly along the circle $\{(x,y)\,|x^2+y^2 = d_i^2\}$ with equal space between consecutive CHs, where $d_i = \frac{r_{i-1}+r_i}{2}$. We assume that the initial battery energy at all CHs is the same. A sensor located in the ith ring is assigned to the nearest CH in the same ring. In the analysis, we assume that a sufficiently large number of CHs are placed in each ring such that the area covered by each CH can be approximated by a small circle, as shown in Figure 1. In the simulations section, we show that this assumption has a negligible impact on network performance.
Remark: Although our model assumes a circular sensing area and a two-tier network structure, the analysis adequately captures the intrinsic interaction between inter- and intra-cluster traffic. The analysis can be extended to handle a non-circular region by covering it with a series of circles, similar to the approach used in cellular networks (in cellular networks, the region is approximately covered by hexagons). A multi-layered organization of sensors, such as the "spine" hierarchy [21], can also be accommodated in our analytical framework. In this case, our analysis provides the optimal CH coverage time for the "base" layers and a sub-optimal coverage time for the whole network. The details of such extensions are beyond the scope of this paper and will be considered in a future work.

2.2 Traffic Model

In the underlying WSN, each sensor generates data at a rate λ (in bits/second). The data are transmitted from the source sensor to its CH, which then forwards the data to the sink, directly or through other CHs. We assume that each sensor has sufficient power to communicate directly with its CH. Furthermore, we assume that the CH depletes its energy at a much faster rate than the sensors it serves. This assumption is justified by the low data rate and duty cycle of commonly used sensors. Accordingly, we focus our attention on energy depletion at CHs. From a strategic point of view, a CH is more critical to the coverage of the network than individual sensors.

Two different routing scenarios are considered:

2.2.1 Shortest-Distance Relay

In this scenario, packets are relayed through the closest CH of the adjacent ring. More specifically, a CH in the ith ring receives traffic originating from its own cluster as well as traffic relayed from CHs in the $(i+1)$th ring, and forwards the combined traffic to the closest CH in the $(i-1)$th ring. Traffic relaying continues hop-by-hop until the sink is reached.

For this scenario, we consider a routing-aware clustering mechanism that balances power consumption at different CHs. Clearly, the *radius profile* of the clusters, $\frac{1}{2}(r_1 - r_0), ..., \frac{1}{2}(r_K - r_{K-1})$, is critical to power consumption at different CHs. For example, reducing $\frac{1}{2}(r_i - r_{i-1})$ results in smaller clusters in the ith ring, which leads to less local traffic from these clusters, shorter transmission distances to subsequent CHs in the $(i-1)$th ring, and a higher number of CHs in the ith ring. Because of the symmetry in the topology and traffic load, the traffic from the CHs in the $(i+1)$th ring will be evenly shared among an increased number of CHs in the ith ring, so the volume of the relayed traffic carried by individual CHs in the ith ring will decrease. All of these factors contribute to a reduced power consumption at the CHs in the ith ring. On the other hand, the reduction in the area of the ith ring must be compensated for by other clusters (e.g., the clusters in the jth ring), because of the fixed number of rings in the system. In an analogous manner, power consumption at CHs in ring j will increase. Therefore, by deliberately adjusting the cluster size in different rings, a more balanced power consumption at different CHs, and hence an increase in the coverage time, is achieved. This is addressed in the routing-aware optimal cluster planning scheme presented in Section 3.2.

2.2.2 *Random Relay*

In this scenario, a CH has two options for data transmission. It may relay the data to the closest CH in the subsequent ring, or it may send it directly to the sink. Let α_i be the fraction of the load that a CH in the ith ring transmits directly to the sink. For a given clustering structure, the *relay probability vector* $(\alpha_1, \ldots, \alpha_K)$ plays a critical role in balancing power consumption at different CHs. For example, increasing α_i will reduce the relayed traffic carried by all CHs in rings 1 to $i-1$. But this comes at the expense of higher power consumption at the CHs in the ith ring, because of the longer transmission distance which, on average, increases from approximately $\frac{1}{2}(r_i - r_{i-2})$ to $\frac{1}{2}(r_i + r_{i-1})$. By deliberately adjusting the relay probabilities at different CHs, a more balanced power consumption at different CHs can be achieved.

In Section 3.3, we propose a clustering-aware optimal random relay scheme that addresses the problem of finding the optimal relay probability vector for a given clustering structure. Specifically, we consider a homogeneous clustering structure, i.e., $r_1 - r_0 = r_2 - r_1 = \ldots = r_K - r_{K-1}$, so that all clusters are expected to cover the same number of sensors. This structure is exactly the "load balanced" clustering presented in [11]. It is highly desirable in practice because of its simplicity. Through numerical examples, we show that the proposed clustering-aware optimal random relay scheme achieves longer expected coverage time compared with pure "load balanced" clustering.

2.3 Wireless Channel Model

We use a Rayleigh fading model to describe the channel between two CHs and also between a CH and the sink. For a transmitter-receiver separation x, the channel gain is given by

$$h(x) = L(d_0)\left(\frac{x}{d_0}\right)^{-n}\xi, \qquad (1)$$

where $L(d_0) \stackrel{\text{def}}{=} \frac{G_t G_r l^2}{16\pi^2 d_0^2}$ is the path loss of the close-in distance d_0, G_t is the antenna gain of the transmitter, G_r is the antenna gain of the receiver, l is the wavelength of the carrier frequency, n is the path loss exponent $(2 \le n \le 6)$, and ξ is a normalized random variable that represents the power gain of the fading. Under the assumption of Rayleigh fading, ξ is exponentially distributed; $\Pr(\xi \le t) = 1 - e^{-t}$.

Because ξ is random, the received signal is also random[2]. Hence, correct reception of a signal can be guaranteed only on a probabilistic basis. In our work, we require that $\Pr\{e_r \ge \tau\} \ge \delta_l$ for reliable reception, where e_r is the energy of the received signal, τ is a predefined energy threshold, and δ_l is the required link reliability. For an end-to-end path of M links that experience independently and identically distributed (i.i.d.) fades, the overall path reliability, i.e., the probability of a successful end-to-end reception, is given by $\delta_p = \delta_l^M$. Therefore, in order to guarantee a path reliability of δ_p, the link reliability δ_l should be at least $\delta_p^{\frac{1}{M}}$. The assumption of i.i.d. link fadings is justified by noting that the distance between consecutive CHs is much larger than the wavelength of the carrier for a system operating at 2.4 GHz, which is a typical value in current standards.

[2]Cost and energy considerations in WSNs prohibit the use of fast power control to combat the fluctuations in channel fading, as typically done in cellular networks.

3. PROPOSED SCHEMES

3.1 Problem Formulation

Let P_i be the average power consumption used for communications by any CH in the ith ring. We adopt the following energy model for P_i:

$$P_i = e_{rx}(\lambda_{oi} + \lambda_{ri}) + e_{tx}(\lambda_{oi} + \lambda_{ri}) + P_{Ti}(\lambda_{oi} + \lambda_{ri}, \Gamma) \quad (2)$$

where e_{rx} is the energy-per-bit consumed in the receive circuit, e_{tx} is the energy-per-bit consumed in the transmit circuit, λ_{oi} is the expected intra-cluster bit rate (in bits/second), λ_{ri} is the expected bit rate of the incoming inter-cluster traffic that is to be relayed by the underlying CH, and $P_{Ti}(.,.)$ is the transmission power expressed as a function of the outgoing bit rate and the employed routing scheme Γ. Note that the terms in the right-hand side of (2) represent, respectively, the power consumption in the receive circuit, the transmit circuit, and the radio interface.

To maximize the expected coverage time, we need to solve the following optimization problem:

$$\text{maximize } \min\left\{\frac{E_1}{P_1}, \frac{E_2}{P_2}, \ldots, \frac{E_K}{P_K}\right\} \qquad (3)$$

where E_i, $i = 1, \ldots, K$, is the initial battery energy for any CH in the ith ring. In practice, CHs are often initialized with identical batteries, i.e., $E_i = E$ for all i. In this case, the optimization problem in (3) becomes equivalent to:

$$\text{minimize } \max\{P_1, P_2, \ldots, P_K\}. \qquad (4)$$

Hereafter, we focus on (4), as this formulation leads to a standard signomial optimization problem, which can be efficiently solved using generalized geometric programming.

3.2 Routing-Aware Optimal Cluster Planning Scheme

In this section, we formulate the optimal cluster organization problem in the context of shortest distance (hop-by-hop) routing. Under this routing scheme, a CH in the ith ring transmits all of its data to the nearest CH in the $(i-1)$th ring. Let x_i be the physical distance between these two CHs. The expected transmission power is given by

$$P_{Ti} = e_{ti}(\lambda_{oi} + \lambda_{ri}) \qquad (5)$$

where e_{ti} is the transmission energy per bit for the underlying CH. Substituting (5) into (2), the expected communication power consumption of any CH at ring i is given by

$$P_i = (e_{rx} + e_{tx} + e_{ti})(\lambda_{oi} + \lambda_{ri}). \qquad (6)$$

Given e_{ti}, the corresponding received energy e_{ri} is given by

$$e_{ri} = e_{ti} L(d_0)\left(\frac{x}{d_0}\right)^{-n}\xi. \qquad (7)$$

The link-reliability requirement can be expressed as

$$\begin{aligned}\delta_l &= \Pr\{e_{ri} \geq \tau\} \\ &= \Pr\left\{\xi \geq \frac{\tau}{e_{ti}L(d_0)}\left(\frac{x_i}{d_0}\right)^n\right\} \\ &= e^{-\frac{\tau x_i^n}{e_{ti}L(d_0)d_0^n}} \end{aligned} \quad (8)$$

Under min-hop routing, the maximum number of links of an end-to-end path is K. Therefore, in order to guarantee the constraint δ_p on the path reliability, the minimum link reliability must be

$$\delta_l = \delta_p^{\frac{1}{K}}. \quad (9)$$

Equating (8) and (9), the minimum transmit energy per bit that satisfies the end-to-end reliability requirement is given by

$$e_{ti} = \frac{-K\tau x_i^n}{L(d_0)d_0^n \log \delta_p}. \quad (10)$$

An approximation that provides an upper bound on the expected coverage time can be obtained by replacing x_i in (10) with a lower bound $x_{i\,\min}$ that is given by:

$$x_{i\,\min} = \begin{cases} \frac{r_1+r_0}{2}, & \text{for } i = 1 \\ \frac{r_i - r_{i-2}}{2}, & \text{for } i = 2,\ldots,K. \end{cases} \quad (11)$$

This lower bound represents the sum of the radii of a cluster in the ith ring and the nearest cluster in the $(i-1)$th ring. It is easy to see that the distance between the CHs of the corresponding two clusters is at least $x_{i\,\min}$.

Let λ_{totali} denote the bit rate of the aggregate traffic that originates from the clusters in rings i through K. Then,

$$\lambda_{totali} = \pi(R^2 - r_{i-1}^2)\rho\lambda, \quad i = 1,\ldots,K. \quad (12)$$

Because relaying is done hop-by-hop, the total traffic load carried by the CHs in the ith ring is equal to the total traffic volume originating from all clusters in rings i to K. Due to the symmetry of the rings and the uniform distribution of sensors, the traffic from the ith ring is evenly distributed among all CHs in that ring. The number of CHs in the ith ring is approximately given by

$$N_i \approx \frac{2\pi r_i}{r_i - r_{i-1}}. \quad (13)$$

The quality of this (and other) approximations is evaluated in Section 4 through a comparison with more realistic simulations.

Accordingly, the average traffic load at any CH in ring i is given by

$$\begin{aligned}\lambda_{oi} + \lambda_{ri} &= \frac{\lambda_{totali}}{N_i} \\ &\approx \frac{(R^2 - r_{i-1}^2)(r_i - r_{i-1})}{2r_i}\rho\lambda. \end{aligned} \quad (14)$$

Substituting (14), (10), and (11) in (6), the expected communication power consumption of any CH in the ith ring can be approximately represented as *signomial functions*[3] of the ring radius profile $\mathbf{r} \stackrel{\text{def}}{=} (r_1, r_2, \ldots, r_K)$. More specifically, they are given by

$$\begin{aligned} P_1 &= \left[e_{rx} + e_{tx} + \frac{K\tau}{-L(d_0)d_0^n \log \delta_p}\left(\frac{r_1 + r_0}{2}\right)^n\right] \times \\ &\quad \frac{(R^2 - r_0^2)(r_1 - r_0)}{2r_1}\rho\lambda, \end{aligned} \quad (15)$$

and

$$\begin{aligned} P_i &= \left[e_{rx} + e_{tx} + \frac{K\tau}{-L(d_0)d_0^n \log \delta_p}\left(\frac{r_i - r_{i-2}}{2}\right)^n\right] \times \\ &\quad \frac{(R^2 - r_{i-1}^2)(r_i - r_{i-1})}{2r_i}\rho\lambda, \text{ for } i = 2,\ldots,K. \end{aligned} \quad (16)$$

Our goal now is to determine the optimal $\mathbf{r}$ that minimizes the average maximum power consumption among all CHs. This optimization problem can be formulated as follows:

$$\begin{cases} \text{minimize}_{\{r_1,\ldots,r_K\}} \{\max\{P_1(\mathbf{r}),\ldots,P_K(\mathbf{r})\}\} \\ s.t. \\ r_0 < r_1 < \ldots < r_K = R \end{cases} \quad (17)$$

where $P_i(\mathbf{r})$, $i = 1,\ldots,K$, are given by (15) and (16).

By introducing the auxiliary variable $t \geq P_i(\mathbf{r})$ for $1 \leq i \leq K$, the optimization problem in (17) can be transformed into the following equivalent form:

$$\begin{cases} \text{minimize}_{\{\mathbf{r},t\}} t \\ s.t. \\ t^{-1}P_i(\mathbf{r}) \leq 1, \quad i = 1,\ldots,K \\ r_{i-1}r_i^{-1} < 1, \quad i = 1,\ldots,K \\ r_K = D. \end{cases} \quad (18)$$

An examination of (18) reveals that its objective function is a *monomial*, the inequality constraints are signomials, and the equality constraint is a monomial of the variables $(\mathbf{r}, t)$ (refer to the appendix for the concepts of monomial, posynomial, and signomial). Therefore, (18) is a signomial optimization problem of the standard form [22]. Its optimal solution can be efficiently found using generalized geometric programming algorithms introduced in [22] and [23].

3.3 Clustering-Aware Optimal Random Relay Scheme

In this section, we consider a "load-balancing" clustering structure and address the optimization of coverage time by determining the optimal relay probabilities at different CHs. Recall that in the current scenario, a CH in the ith ring relays its traffic to the closest CH in the $(i-1)$th ring with probability $1-\alpha_i$, and with probability α_i, it transmits directly to the sink. Under "load-balanced" clustering, homogeneous clusters are formed, so that each of these clusters is expected to contain the same number of sensors. In this case, $r_i = r_0 + i\frac{R-r_0}{K}$, for $i = 1,\ldots,K$.

Under random relay routing with $i \geq 2$, the average power consumed to transmit data from any CH in the ith ring is

$$P_{Ti} = e_{tri}(1-\alpha_i)(\lambda_{oi} + \lambda_{ri}) + e_{tdi}\alpha_i(\lambda_{oi} + \lambda_{ri}) \quad (19)$$

where e_{tri} and e_{tdi} are the transmission energy per bit for relaying traffic to the nearest CH in ring $i-1$ and for direct transmission to the sink, respectively. Following a similar development to the one in Section 3.2, e_{tri} and e_{tdi} are derived as follows:

$$e_{tri} = \frac{-K\tau x_i^n}{L(d_0)d_0^n \log \delta_p} \quad (20)$$

[3] See the appendix for the definition of signomial functions.

and

$$e_{tdi} = \frac{-(K-i+1)\tau d_i^n}{L(d_0) d_0^n \log \delta_p} \quad (21)$$

where x_i and d_i are the distances for the one-hop relay and for the direct CH-to-sink transmission, respectively. In (21), the factor $(K-i+1)$ is used instead of K in (20) and (10) to accommodate a worst-case link reliability requirement. Recall that in deriving (10), we split the end-to-end path reliability δ_p among K links, providing a conservative estimate of the link reliability for each of the K hops. In the case of the random relay scheme, the traffic that is relayed to a CH in the ith ring from outer rings may have traversed from one to $K-i$ hops before reaching the ith ring. So if this traffic is to be transmitted directly from the ith ring to the sink, its maximum hop count would be $K-i+1$, which explains the appearance of this factor in (21).

The CH-to-CH distance x_i in (20) is lower bounded by

$$x_{i\,\min} = \frac{R-r_0}{K} \quad (22)$$

which represents the sum of the radius of a cluster in the ith ring and the radius of the closest cluster in the $(i-1)$th ring (note that under the random relay strategy, rings have the same "thickness," so $r_i - r_{i-1} = (R-r_0)/K$ for all $i = 2, 3, \ldots, K$).

For the first ring, there is no difference between relaying to the next CH and directly communicating with the sink. Therefore,

$$P_{T1} = e_{td1}(\lambda_{o1} + \lambda_{r1}) \quad (23)$$

where e_{td1} is given by

$$e_{td1} = \frac{K\tau d_1^n}{-L(d_0) d_0^n \log \delta_p}. \quad (24)$$

The traffic load coming from any CH in the ith ring, i.e., $\lambda_{oi}+\lambda_{ri}$, is derived analytically as follows. Let λ_{totali} denote the total traffic load that originates from all clusters in rings i to K. For the Kth ring,

$$\lambda_{totalK} = \pi(R^2 - r_{K-1}^2)\rho\lambda. \quad (25)$$

Then, the traffic load relayed from the Kth ring to the $(K-1)$th ring is simply given by

$$(1-\alpha_K)\lambda_{totalK} = \left[\pi R^2(1-\alpha_K) - \pi r_{K-1}^2(1-\alpha_K)\right]\rho\lambda. \quad (26)$$

For the $(K-1)$th ring,

$$\begin{aligned}\lambda_{totalK-1} &= \left[\pi(r_{K-1}^2 - r_{K-2}^2)\right]\rho\lambda + (1-\alpha_K)\lambda_{totalK} \\ &= \left[\pi R^2(1-\alpha_K) + \pi r_{K-1}^2\alpha_K - \pi r_{K-2}^2\right]\rho\lambda. \quad (27)\end{aligned}$$

Similarly, one can easily derive $\lambda_{totalK-2}, \ldots, \lambda_{total1}$. An examination of these equations shows that they can be expressed as

$$\lambda_{totali} = \left[\pi \sum_{j=i}^{K} r_j^2 \alpha_{j+1} \prod_{m=i+1}^{j} (1-\alpha_m) - \pi r_{i-1}^2\right]\rho\lambda, \quad \text{for } i = 1, \ldots, K \quad (28)$$

where we take α_{K+1} to be 1.

Following a similar analysis to the one used to arrive at (14), the total traffic load produced by any CH in the ith ring, including intra-cluster as well as relayed traffic, can be written as

$$\begin{aligned}\lambda_{oi} + \lambda_{ri} &= \lambda_{totali}\frac{R-r_0}{2\pi K r_i} \\ &= \left[\pi \sum_{j=i}^{K} r_j^2 \alpha_{j+1} \prod_{m=i+1}^{j} (1-\alpha_m) - \pi r_{i-1}^2\right] \times \\ &\rho\lambda\frac{R-r_0}{2\pi K r_i}. \quad (29)\end{aligned}$$

Substituting (19), (20), (21), and (29) into (2), the expected communication power consumption of any CH at ring i is given by

$$\begin{cases} P_1 = \left[e_{rx} + e_{tx} + \frac{K\tau}{-L(d_0)\log\delta_p}\left(\frac{d_1}{d_0}\right)^n\right] \times \\ \left[\pi \sum_{j=1}^{K} r_j^2 \alpha_{j+1} \prod_{m=2}^{j}(1-\alpha_m) - \pi r_0^2\right]\rho\lambda\frac{R-r_0}{2\pi K r_1} \\ P_i = \left[e_{rx} + e_{tx} + \alpha_i\frac{(K-i+1)\tau}{-L(d_0)\log\delta_p}\left(\frac{d_i}{d_0}\right)^n + \right. \\ \left.(1-\alpha_i)\frac{K\tau}{-L(d_0)\log\delta_p}\left(\frac{D-r_0}{Kd_0}\right)^n\right] \times \\ \left[\pi \sum_{j=i}^{K} r_j^2 \alpha_{j+1} \prod_{m=i+1}^{j}(1-\alpha_m) - \pi r_{i-1}^2\right] \times \\ \rho\lambda\frac{R-r_0}{2\pi K r_i}, \quad i = 2, \ldots, K \end{cases} \quad (30)$$

where $r_i = r_0 + i\frac{R-r_0}{K}$ and $d_i = r_i - \frac{R-r_0}{2K}$.

¿From (30), it is clear that for a given radius profile $(r_0, r_2, \ldots, r_K)$, the expected power consumption at different CHs can be expressed as signomial functions of the probabilities $(\alpha_1, \ldots, \alpha_K)$. Our goal is to determine the optimal values for these probabilities that maximize the expected coverage time (equivalently, minimize the maximum expected power consumption at a CH). More specifically, this optimization problem can be formulated as follows:

$$\begin{cases} \text{minimize}_{\{\alpha_1,\ldots,\alpha_K\}} \max\{P_1, \ldots, P_K\} \\ s.t. \\ \alpha_1 = 1 \\ 0 \le \alpha_i \le 1, \quad i = 2, \ldots, K \end{cases} \quad (31)$$

where P_i's are given in (30).

By introducing the auxiliary variable t, (31) can be transformed into the following equivalent optimization problem:

$$\begin{cases} \text{minimize}_{\{\alpha_1,\alpha_2,\ldots,\alpha_K,t\}} t \\ s.t. \\ t^{-1}P_i \le 1, \quad i = 1, \ldots, K \\ \alpha_i \le 1, \quad i = 2, \ldots, K \\ 1-\alpha_i \le 1, \quad i = 2, \ldots, K \\ \alpha_1 = 1. \end{cases} \quad (32)$$

An examination of (32) and (30) shows that the objective function is a monomial, the inequality constraints are signomials, and the equality constraint is a monomial in the variables $(\alpha_1, \ldots, \alpha_K, t)$. Therefore, (32) is also a generalized geometric programming problem in a standard form, which can be efficiently solved by the GGP algorithms introduced in [22] and [23] (also, refer to the appendix).

Remark: As verified in Section 4, in most cases, the objective functions in (32) and (18) are minimized when power consumptions at different CHs are equalized. This is because if there is a CH with power P_i that is larger than the power consumption of other CHs, then P_i can always be lowered without violating the constraints by decreasing r_i in (18) or α_i in (32), leading to an increase in the power consumption of some other CHs. As a result, the maximum power consumption will be minimized when a balance is reached across all CHs.

4. NUMERICAL RESULTS AND SIMULATIONS

In this section, we study the performance of the proposed optimal cluster planning and optimal random relay schemes and compare them with a "load-balanced" clustering approach [11] that uses hop-by-hop traffic relay between CHs in consecutive rings. The analysis conducted in the previous sections was based on certain simplifying assumptions (e.g., circular clusters, lower bounds on CH-to-CH distances, etc.). To validate the adequacy of our analytical results, we contrast them with simulations conducted in a more realistic setting (explained below). For the two proposed schemes, we use the analytical results to compute the optimal radius profile $\mathbf{r}$ and optimal relay probabilities $\alpha_1, \ldots, \alpha_K$. We use these optimal values to drive the simulations of the two proposed schemes. Our main performance metric is the maximum expected power consumption of a CH, $P_{\max} \stackrel{\text{def}}{=} \max\{P_1, \ldots, P_K\}$. The smaller the value of $P_{\max}$, the longer is the coverage time. We set the radius of the circular sensing region to $R = 200m$, where m is an arbitrarily chosen distance unit (e.g., meter). Sensors are uniformly distributed throughout this region at density $\rho = 1/m^2$, i.e., the number of sensors in any area S follows a spatial Poisson distribution with parameter ρS. The number of CHs in both the analysis and the simulations is set to $\sum_{i=1}^{K} N_i$, where N_i is obtained from (13) and K is given. The location of these CHs is also taken to be the same for the analysis and the simulations. However, in the simulations, clusters are not necessarily circular, and the notion of rings is no longer applicable. Instead, each sensor in a given simulation run is assigned to the nearest CH. As a result, two CHs that have the same distance to the sink may have different cluster sizes. Each sensor generates data according to a Poisson process of rate $\lambda = 10$ bits/second[4]. Because of the randomness in the traffic and node distribution, the powers consumed by different CHs in the same ring may be different in the simulations. In this case, $P_{\max}$ is taken as the maximum of $P_{avg,1}, \ldots, P_{avg,K}$, where $P_{avg,i}$ is the average power of a CH in the ith ring. We take $r_0 = 10m$, $G_t = G_r = 1$, $\tau = 10^{-17}$ Joules, and $\delta_p = 0.99$ (the required end-to-end path reliability).

Figures 2 and 3 depict $P_{\max}$ versus the number of rings (K) for two path loss factors: $n = 2$ and $n = 4$, respectively. It is observed that the gap between the (approximate) analytical results and the simulations is reasonably small for all examined schemes, with the simulation results being slightly more conservative than the analysis. The disparity between the two is attributed in part to the approximate nature of the analysis and in part to the randomness in the packet generation process and the distribution of sensors within a CH. When $n = 2$, both the optimal cluster planning and the optimal random relay schemes result in significantly longer coverage times (smaller $P_{\max}$ values) than the "load-balanced" clustering scheme. This improvement is attributed to the power-balancing philosophy used in the two proposed schemes. For $n = 4$ (Figure 3), the optimal cluster planning scheme maintains its advantage, but the optimal random relay scheme is shown to achieve only limited power efficiency over "load-balanced" clustering. This can be ex-

[4]The choice of the traffic model has no impact on the relative performance of the investigated schemes. For this reason, we opted for a simple model.

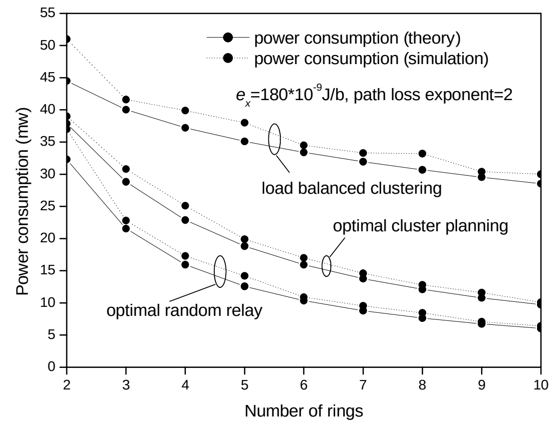

Figure 2: $P_{\max}$ vs. number of rings ($e_x = 180$ nJ/bit, $n = 2$).

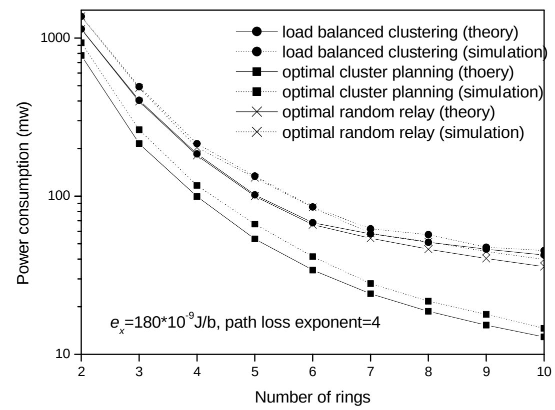

Figure 3: $P_{\max}$ vs. number of rings ($e_x = 180$ nJoule/bit, $n = 4$).

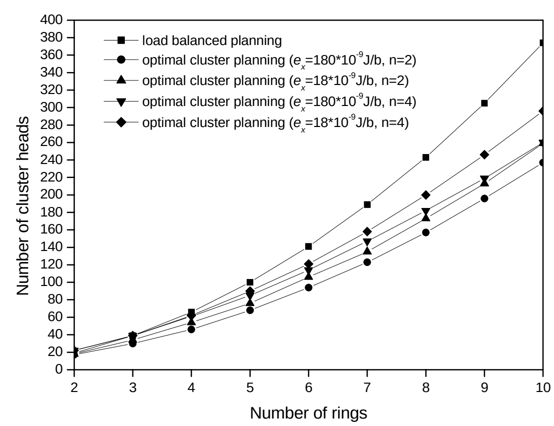

Figure 4: Number of clusters vs. number of rings (K).

plained by noting that for a large path-loss exponent n, the total power consumption is dominated by the transmission power (P_{Ti}), which is highly nonlinear in the transmission distance. As a result, for the random relay scheme, only a small portion of the traffic at each CH (in our simulations, less than 5%) is transmitted directly to the sink; the rest is sent hop-by-hop. For traffic sent hop-by-hop, the random relay and the "load-balanced" clustering schemes have comparable $P_{\max}$ values (they both use equal-size clusters).

Figure 4 depicts the total number of formed clusters ($\sum_{i=1}^{K} N_i$) versus the number of rings (K) for the optimal cluster planning and the "load-balanced" schemes. In addition to achieving a lower $P_{\max}$ value (longer coverage time), optimal cluster planning also results in a smaller number of clusters (and hence, reduced network-management overhead). The reduction in the number of clusters comes from the improved energy utilization of under-drained CHs, i.e., in order to balance the power consumption of different CHs, an under-drained CH tends to carry more intra-cluster traffic, hence expanding the size of the cluster and reducing the number of clusters required to cover the sensing region.

In Figure 5, we study the effect of the transmit-plus-receive per-bit *circuit* energy $e_x \overset{\text{def}}{=} e_{tx} + e_{rx}$ on the performance of the two proposed schemes, using $n = 2$ and $K = 10$. Interestingly, when e_x is small (the circuit is more energy-efficient), optimal cluster planning achieves better coverage time (smaller $P_{\max}$) than optimal random relay. However, as e_x increases, the relative difference between the two schemes shrinks, and eventually optimal random relay becomes superior in terms of coverage-time performance to optimal cluster planning.

This phenomenon can be explained as follows. When e_x is small, the total power consumption at a CH is dominated by the transmission power (P_{Ti}), which is minimized when the data are forwarded hop-by-hop using shortest-distance routing. Because optimal cluster planning relies solely on shortest-distance routing, whereas optimal random relay sometimes uses direct CH-to-sink communication, the former scheme achieves a lower $P_{\max}$ value. As e_x increases, circuit power becomes more significant, and multi-hop (shortest-distance) routes become less energy-efficient. In this case, direct CH-to-sink communication becomes more attractive, giving optimal random relay an advantage over optimal cluster planning.

In Figures 6 and 7, we study via simulations the effect of balancing the power across different rings. As indicated earlier, for the optimal cluster planning and optimal random relay schemes, the radius profile and the relay probabilities are obtained from the analysis and used in the simulations. We measure the accuracy in the power balance by $\eta \overset{\text{def}}{=} \frac{Std(P_{avg,1},\ldots,P_{avg,K})}{Avg(P_{avg,1},\ldots,P_{avg,K})}$. The smaller the value of η, the more balanced is power consumption across different CHs (and the larger is the coverage time). The figures indicate that in most cases, the analysis-based optimization of the radius profile and relay probabilities leads to a small η (e.g., less than 0.1). However, Figure 7 shows that for a small K and $n = 4$, the optimal random relay scheme exhibits a relatively large η (comparable with the value of η for "load balanced" clustering). This can be explained by noting that for a small K, the length of each CH-to-CH hop is considerably larger than the distance between the sink and a CH in the first ring. Under a highly nonlinear channel attenu-

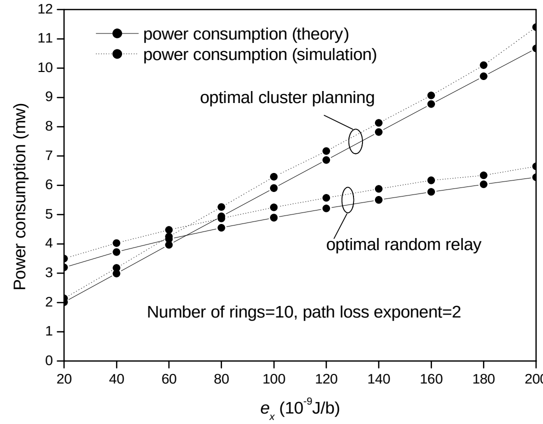

Figure 5: $P_{\max}$ vs. circuit energy efficiency e_x ($K = 10$, $n = 2$).

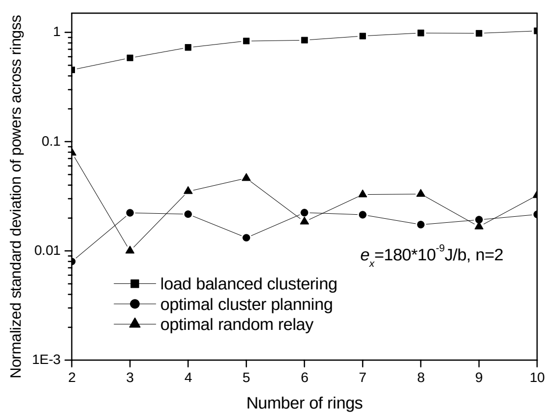

Figure 6: Normalized standard deviation of power consumption vs. number of rings ($e_x = 180$nJ/b, $n = 2$).

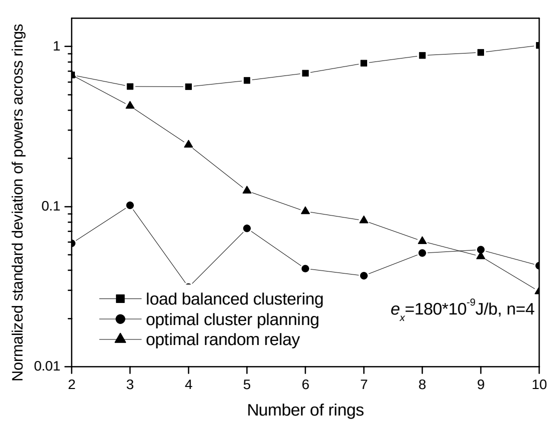

Figure 7: Normalized standard deviation of power consumption vs. number of rings ($e_x = 180$nJ/b, $n = 4$).

ation model ($n = 4$), even if $\alpha_i = 0$ (i.e., no traffic is sent directly to the sink), the power consumption for CH-to-CH relaying is still much larger than the power consumption of a CH in the first ring. Consequently, no power balance can be reached in this scenario. As we increase K, the length of each relay hop decreases, so the power tradeoff between relay and direct transmission becomes dominant in the optimization, leading to a better power balance.

5. CONCLUSIONS

We considered the problem of coverage-time optimization by balancing power consumption at different CHs in a clustered WSN. Our study demonstrates the significance of simultaneously accounting for the impacts of intra- and inter-cluster traffic in the design of routing and clustering strategies. Two mechanisms for balancing power consumption were studied: the (routing-aware) optimal cluster planning and the (clustering-aware) optimal random relay. Under the assumptions of circular sensing area and clusters, the control parameters in both mechanisms (radius profile and relay probabilities) were optimized with respect to the maximum power consumption of a CH. The optimization problems were formulated as signomial optimizations, which were efficiently solved using generalized geometric programming. For tractability purposes, our analysis is necessarily approximate, as it relies on several simplifying assumptions. Simulations were conducted to verify the adequacy of our analysis and demonstrate the substantial benefits of the two proposed schemes in terms of prolonging the coverage time of the network.

The definition of coverage time used in our work is somewhat conservative, and mainly applies to application scenarios with stringent coverage requirements. Our future efforts will focus on other, less stringent (and more general) definitions of coverage time, including the battery drainage for $x\%$ of CHs or the time until the network partitions. We will also consider extending the analysis to hierarchically clustered WSNs (e.g., the "spine" hierarchy).

Acknowledgements

This work was supported by NSF under grants ANI-0095626, ANI-0313234, and ANI-0325979, and by the Center for Low Power Electronics (CLPE) at the University of Arizona. CLPE is supported by NSF (grant # EEC-9523338), the State of Arizona, and a consortium of industrial partners. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

6. REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, pp. 102-114, Aug. 2002.

[2] J. M. Kahn, R. H. Katz, and K. S. Pister, "Next century challenges: mobile networking for smart dust," in *Proc. ACM/IEEE MobiCom '99 Conf.*, pp. 271-278, 1999.

[3] D. J. Baker and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1694-1701, Nov. 1981.

[4] A. Ephremides, J. E. Wieselthier, and D. J. Baker, "A design concept for reliable mobile radio networks with frequency hopping signaling," *Proceedings of the IEEE*, vol. 75, no. 1, pp. 56-73, 1987.

[5] C. R. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1265-1275, Sep. 1997.

[6] M. Gerla and J.T.C. Tsai, "Multicluster, mobile, multimedia radio network," *ACM/Baltzer Journal of Wireless Networks*, vol. 1, no. 3, pp. 255-265, 1995.

[7] J. Wu and H. L. Li, "On calculating connected dominating set for efficient routing in ad hoc wireless networks," *Proc. of the 3rd ACM International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pp. 7-14, 1999.

[8] F. Kuhn, T. Moscibroda, and R. Wattenhofer, "Initializing newly deployed ad hoc and sensor networks," *Proc. ACM MobiCom'04 Conference*, Philadelphia, Sep. 2004.

[9] P. J. Wan, K. M. Alzoubi, and O. Frieder, "Distributed construction of connected dominating set in wireless ad hoc networks," *Proc. IEEE INFOCOM Conference*, New York, June 23-27, 2002.

[10] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660-670, Oct. 2002.

[11] G. Gupta and M. Younis, "Load-balanced clustering of wireless sensor networks," in *Proc. IEEE International Conference on Communications (ICC'03)*, vol. 3, pp. 1848-1852, May 2003.

[12] A. D. Amis and R. Prakash, "Load-balancing clusters in wireless ad hoc networks," in *Proc. 3rd IEEE Symposium on Application-Specific Systems and Software Engineering Technology*, pp. 25-32, Mar. 2000.

[13] A. D. Amis, R. Prakash, T. H. P. Vuong, and D. T. Huynh, "Max-min d-cluster formation in wireless ad hoc networks," in *Proc. IEEE INFOCOM 2000 Conf.*, vol. 1, pp. 32-41, Mar. 2000.

[14] C. F. Chiasserini, I. Chlamtac, P. Monti, and A. Nucci, "An energy efficient method for nodes assignment in cluster-based ad hoc networks," *ACM/Kluwer Wireless Networks Journal (WINET)*, vol. 10, no. 3, pp. 223-231, May 2004.

[15] S. Bandyopadhyay, E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. IEEE INFOCOM 2003 Conference*, vol. 3, pp. 1713-1723, Mar. 2003.

[16] S. Bandyopadhyay and E. J. Coyle, "Minimizing communication costs in hierarchically-clustered networks of wireless sensors," *Elsevier Journal of Computer Networks*, vol. 44, no. 1, pp. 1-16, Jan. 2004.

[17] T. Moscibroda and R. Wattenhofer, "Maximizing the lifetime of dominating sets, " *Proc. WMAN'05 Conference*, 2005.

[18] J. Pan, Y. T. Hou, L. Cai, Y. Shi, and S. X. Shen, "Topology control for wireless sensor networks," in *Proc. ACM MobiCom 2003 Conf.*, pp. 286-299, 2003.

[19] S. J. Park, R. Vedantham, R. Sivakumar, and I. F. Akyildiz, "A scalable approach for reliable downstream data delivery in wireless sensor networks," in *Proc. ACM MobiHoc 2004 Conf.*, pp. 78-89, May 2004.
[20] C. F. Chiasserini and M. Garetto, "Modeling the performance of wireless sensor networks," in *Proc. IEEE INFOCOM 2004 Conf.*, Mar. 2004.
[21] B. Das, R. Sivakumar, and V. Bharghavan, "Routing in ad hoc networks using a spine," in *Proc. IEEE International Conference on Computer Communications and Networks*, pp. 34-39, Sept. 1997.
[22] J. G. Ecker, "Geometric programming: methods, computations and applications," *Society of Industrial and Applied Mathematics, SIAM Review*, vol. 22, no. 3, pp. 338-362, July 1980.
[23] E. L. Peterson, "Geometric programming," *Society of Industrial and Applied Mathematics, SIAM Review*, vol. 18, no. 1, pp. 1-51, Jan. 1976.
[24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, U.K., 2004.

APPENDIX

A. GENERALIZED GEOMETRIC PROGRAMMING

A function h is a *monomial* in the variables $x_1, x_2, \ldots, x_n$ if it can be written as $h(x_1, \ldots, x_n) x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$ for any real-valued exponents $a_1, \ldots, a_n$. Furthermore, a function f is a *posynomial* in the variables $x_1, x_2, \ldots, x_n$ if it can be written as

$$f(x_1, \ldots, x_n) = \sum_{j=1}^{L} c_j g_j(x_1, \ldots, x_n) \tag{33}$$

where for $j = 1, \ldots, L$, $c_j \geq 0$ and g_j is a monomial in $x_1, x_2, \ldots, x_n$.

Let $\mathbf{x} \stackrel{\text{def}}{=} (x_1, x_2, \ldots, x_n)$ be a vector of n variables and let M_1 and M_2 be two positive integers. A standard geometric program is an optimization problem of the form:

$$\begin{cases} \min f_0(\mathbf{x}) \\ s.t. \\ f_i(\mathbf{x}) \leq 1, \quad i = 1, \ldots, M_1 \\ h_l(\mathbf{x}) = 1, \quad l = 1, \ldots, M_2 \end{cases} \tag{34}$$

where $f_0, f_1, \ldots, f_{M_1}$ are posynomials in $\mathbf{x}$ and $h_1, \ldots, h_{M_2}$ are monomials in $\mathbf{x}$.

A geometric program in the standard form is not a convex optimization problem. However, with the change of variables $y_i \stackrel{\text{def}}{=} \log x_i$ and $b_i \stackrel{\text{def}}{=} \log c_i$, it can be transformed into the following convex form:

$$\begin{cases} \min \left\{ p_0(\mathbf{y}) \stackrel{\text{def}}{=} \log \sum_j \exp\left(\mathbf{a}_{0j}^T \mathbf{y} + b_0\right) \right\} \\ s.t. \\ p_i(\mathbf{y}) \stackrel{\text{def}}{=} \log \sum_j \exp\left(\mathbf{a}_{ij}^T \mathbf{y} + b_i\right) \leq 0, \quad i = 1, \ldots, M_1 \\ q_l(\mathbf{y}) \stackrel{\text{def}}{=} \mathbf{a}_l^T \mathbf{y} + b_l = 0, \quad l = 1, \ldots, M_2 \end{cases} \tag{35}$$

where $\mathbf{a}_{ij} = (a_{ij1}, a_{ij2}, \ldots, a_{ijn})^T \in \mathbf{R}^n$ is the exponent vector of the jth monomial in the ith posynomial and $\mathbf{y} \stackrel{\text{def}}{=} (y_1, \ldots, y_n)^T$ is the optimization variable. The logarithm of a sum of exponentials is a convex function. Thus, (35) is a convex optimization problem that can be efficiently solved using numerical algorithms such as the interior point method [24].

A signomial is a more generalized form of a posynomial, whereby the coefficients c_j, $j = 1, \ldots, L$, can have any real values. If in (34) the constraints consist of signomials, the formulation is called a *signomial program* or *generalized geometric programming.*

Any signomial program can be transformed into an equivalent program of the form

$$\begin{cases} \min g_0(\mathbf{x}) \\ s.t. \\ g_k(\mathbf{x}) \leq 1, \quad k = 1, \ldots, p \\ g_k(\mathbf{x}) \geq 1, \quad k = p+1, \ldots, q \end{cases} \tag{36}$$

where $g_k(\mathbf{x})$ is a posynomial for $k = 0, 1, \ldots, q$. The form (36) is called a reversed posynomial program.

One approach for solving signomial problems is to "condense" the posynomial in each reversed constraint (i.e., approximate the sum of monomials by using their geometric average, leading to another monomial) and obtain a posynomial program that approximates the original signomial program. Upon solving the posynomial program by any convex optimization algorithm, the solution is used to generate a better approximation. For example, suppose a program $\mathbf{S}$ of the form (36) contains a single reversed constraint

$$g_l(\mathbf{x}) \geq 1. \tag{37}$$

Let $\dot{g}_l(\mathbf{x})$ be the monomial obtained by condensing g_l with an arbitrary set of weights ϵ using the arithmetic-geometric mean inequality. Let $\dot{\mathbf{S}}$ denote the program obtained from $\mathbf{S}$ where (37) is replaced by

$$\dot{g}_l(\mathbf{x})^{-1} \leq 1. \tag{38}$$

Since $\dot{g}_l(\mathbf{x})$ is a monomial, (38) is a standard posynomial constraint and $\dot{\mathbf{S}}$ is a posynomial program that approximates the signomial program $\mathbf{S}$. Moreover, the arithmetic-geometric inequality implies that $\dot{g}_l(\mathbf{x}) \leq g_l(\mathbf{x})$. Thus, if $\mathbf{x}$ is feasible for $\dot{\mathbf{S}}$, then it is feasible for $\mathbf{S}$. The minimum value for $\dot{\mathbf{S}}$, $M(\dot{\mathbf{S}})$, is an upper bound on the minimum value for $\mathbf{S}$, $M(\mathbf{S})$. Suppose that $\dot{\mathbf{x}}$ is optimal for $\dot{\mathbf{S}}$. Define a new set of weights

$$\epsilon_i = \frac{f_i(\dot{\mathbf{x}})}{g_l(\dot{\mathbf{x}})}. \tag{39}$$

Using these weights, one can define a new condensed posynomial $\ddot{g}_l(\mathbf{x})$ and form the program $\ddot{\mathbf{S}}$ where $\ddot{g}_l(\mathbf{x}) \geq 1$ replaces $g_l(\mathbf{x}) \geq 1$ in $\mathbf{S}$. Since $\ddot{g}_l(\dot{\mathbf{x}}) = g_l(\dot{\mathbf{x}})$ and $\dot{\mathbf{x}}$ is feasible for $\mathbf{S}$, it follows that $\dot{\mathbf{x}}$ is feasible for $\ddot{\mathbf{S}}$. The minimum value for $\ddot{\mathbf{S}}$, $M(\ddot{\mathbf{S}})$, therefore satisfies

$$M(\mathbf{S}) \leq M(\ddot{\mathbf{S}}) \leq M(\dot{\mathbf{S}}). \tag{40}$$

This defines an iterative process for generating a sequence of posynomial programs whose minimum values are monotonically decreasing upper bounds of the desired minimum value for $\mathbf{S}$.

The detailed algorithmic description of signomial programming is out of the scope of this work. A comprehensive survey on algorithms for generalized geometric programming is given in [22].