On the Fundamental Capacity and Lifetime Limits of Energy-Constrained Wireless Sensor Networks

Zhihua Hu, Baochun Li

***Abstract*— Energy constraints on sensor nodes pose significant challenges towards extending operational lifetimes and sustainable capacities of wireless sensor networks. In this paper, we seek to answer two fundamental questions with respect to energy-constrained sensor networks. First, what is the operational lifetime of a particular wireless sensor network under the control of optimal power management schemes? With an adequate definition of operational lifetimes, our asymptotic analysis shows that, with fixed node densities, operational lifetime of sensor networks decreases in the order of $1/n$ as the number of initially deployed nodes n grows. Second, what is the impact of constrained energy levels on the maximum sustainable throughput in sensor networks? Even with renewable energy sources on each of the sensors (e.g., solar energy sources), our analysis concludes that the maximum sustainable throughput in energy-constrained sensor networks scales worse than the capacity based on interference among concurrent transmissions as long as the physical network size grows with n in the order greater than $\log n$. In this case, when the number of nodes is sufficiently high, the energy-constrained network capacity dominates.**

I. Introduction

A fundamental limitation of sensor networks is the constrained energy source at each node ($<$ 0.5 Ah, 1.2V [4]), since most of sensors are micro-electronic devices. During signal propagation, the signal decays as $r^{-\alpha}$ with transmission range r, where α is the loss exponent of the signal [9]. The limited power and signal loss during propagation impose fundamental constraints on the operational lifetime of the sensor network, and other performance issues such as the capacity of data transmissions. In most cases, it is impossible to replenish energy levels in the sensor nodes. In this case, the initial energy levels in the sensor nodes and ongoing energy consumption rates directly affect the operational lifetime and the data transmission capacity of the sensor network. It is therefore evident that effective power management mechanisms are of utmost importance in sensor network designs.

Power management in multi-hop wireless networks such as wireless sensor networks involves energy-aware topology control, scheduling and routing mechanisms. These problems are non-trivial and extensively studied in previous work [1], [5], [15]. In this paper, we take a different angle when we examine the problem of extending the operational lifetime of wireless sensor networks. Given a set of network characteristics and definitions, we seek to answer the following fundamental question: *What is the operational lifetime of a particular wireless sensor network under the control of optimal power management schemes?* An answer to this question leads to insights on the fundamental limits with respect to the performance gains using *any* energy-aware algorithms and protocols. As well, additional insights on the scalability of wireless sensor networks with respect to their energy costs may be derived when we study the relationships between the network lifetimes and sizes.

Zhihua Hu and Baochun Li are affiliated with the Department of Electrical and Computer Engineering, University of Toronto. Their email addresses are {*frank,bli*}*@iqua.ece.toronto.edu*.

Studies on the scalability of sensor networks may lead to surprising results that must be well understood at the time of network deployment. Naturally, a sensor network that fails to function towards the end of its mission should not be deployed, or should be replenished by deploying additional nodes before functional failures. With adequate analysis, we may observe that the network lifetime after its initial deployment may not be arbitrarily extended by simply increasing the number of nodes initially deployed. Before communication failures due to energy costs, provisions must be made to replenish the network by adding additional nodes on the fly after its initial deployment.

Towards extending the lifetime, strategies with respect to such *network replenishment* due to sensor energy costs have never been studied in previous work. However, we argue that these are critical to the lifetime of sensor networks. A simple strategy may be that, a minimum number of nodes is deployed initially, with new nodes subsequently added to the network according to certain schedules. However, the optimal timing, location and size of node additions are still unknown. Theoretical studies on influential factors with respect to sensor network lifetime may lead to insights towards optimal network replenishment strategies.

Addressing these fundamental questions on sensor network lifetime, our original contributions in this paper are as follows. *First*, we rigorously define the concept of *operational lifetime* of sensor networks. After such *lifetime* expires, a certain percentage of data transmissions fails. Though more complex, such a concept of lifetime is more relevant than definitions in previous studies, e.g., the time elapsed until the last sensor node fails [14]. We show that network fails to function with respect to transmissions long before the failure of the last node. *Second*, we develop the lower and upper bounds of operational lifetime using a stochastic model and a cut-based methodology. Our asymptotic analysis shows that, for fixed network sizes, operational lifetime decreases in the order of $1/\sqrt{n}$ as the number of initially deployed nodes n grows. For fixed node densities, the lifetime decreases in the order of $1/n$. Our analysis also shows that the operational lifetime of the network is shorter than the average lifetime of individual nodes by a certain factor, which supports our definition of operational lifetime. *Finally*, we examine the impact of constrained energy levels on the maximum sustainable throughput in sensor networks. For sensor networks with renewable energy sources (e.g., solar energy sources), our analysis shows that the maximum sustainable throughput in energy-constrained sensor networks scales worse than the capacity predicted based on interference among concurrent transmissions, if the physical network size grows with n in the order greater than $\log n$. In this case, when the number of nodes is sufficiently high, the energy-constrained network capacity dominates. We believe that the effects of energy constraints on the operational lifetime and the capacity of wireless sensor networks are still largely unchartered territories, as there exists no previous work seeking to answer these questions analytically to the best of our knowledge.

The remainder of the paper is structured as follows. Sec. II-A summarizes the main contribution of this paper in the form of Theorem II-A 1 and 2. Sec. II-B to Sec. II-D develop the definitions and lemmas required towards the final proof of the aforementioned theorems. The two theorems are proved in Sec. II-E and Sec. II-F. Sec. III discusses related work. Finally, Sec. IV concludes the paper.

II. Energy-constrained Lifetime and Capacity in Sensor Networks

A. Summary of Results

The primary functionality of wireless sensor networks is to sense the environment and transmit the acquired information for further processing. As a result of constrained energy levels, sensors will eventually fail. However, intuitively, a network may fail to continuously support data transmissions — its primary functionality — long before the last sensor node fails (such intuition is shown to be correct later in this paper). This will occur when the number of failed nodes in the network reaches a certain critical threshold. Therefore, the *operational lifetime* of a network should be defined such that, after such lifetime expires, a certain percentage of data transmissions fail.

Let ϵ be a real number that satisfies $0 < \epsilon < 1$, we define the operational lifetime of a wireless sensor network as follows (detailed derivations that motivate such a definition are postponed to Sec. II-D).

Definition II-A 1. The *operational lifetime* of a network is the expected time after which at least $100(1-\epsilon^2)\%$ data transmissions fail.

The understanding of the asymptotic behavior of operational lifetimes is essential to the study of sensor network feasibility: whether or not a sensor network can function till the end of its mission. If a sensor network is proved to be infeasible, either the network should not be deployed, or a *network replenishment strategy* has to be devised. The replenishment strategy may propose to add additional nodes — and thus to add additional energy — to the network, in order to ensure that the network will complete its mission successfully.

In this paper, we systematically study the lower and upper bounds of operational lifetime based on a stochastic model, and then identify its influential factors. Let $b_\epsilon = \frac{1}{\epsilon(2-\epsilon)}$ [1]. Our key results with respect to the operational lifetime are established in the following theorem.

Theorem II-A 1.

(1) For fixed network sizes, the operational lifetime of a wireless sensor network decreases in the order of $1/\sqrt{n}$ as the number of nodes n grows.

(2) For fixed node densities, the operational lifetime of a wireless sensor network decreases in the order of $1/n$.

(3) The operational lifetime of a wireless sensor network is smaller than the average lifetime of individual nodes by a factor of $b_\epsilon/\sqrt{n}$ for fixed network sizes and b_ϵ/n for fixed node densities.

Before showing the proof of this theorem in Sec. II-E, we first illustrate the implications and significance of this theorem. Since $\sum_i \frac{1}{\sqrt{n_i}} > \frac{1}{\sqrt{\sum_i n_i}}$ and $\sum_i \frac{1}{n_i} > \frac{1}{\sum_i n_i}$, Theorem II-A 1 shows that, a good network replenishment strategy is to replenish the network by adding additional batches of sensor nodes in subsequent stages, and these batches should be organized so that the sizes of different stages are as small as possible.

However, there exist several constraints to the smallest batch. These constraints include: (1) The requirement of minimum coverage for sensing purposes (both in terms of node density and network size); (2) The deployment overhead incurred in addition to the cost of sensors; and (3) the limited deployment window due to realistic causes (e.g., enemy positions in battlefields or weather conditions). Theorem II-A 1 may be used to identify the optimal network replenishment strategy under such constraints.

For sensor networks that rely on renewable energy sources such as solar energy, the maximum amount of data that can be transmitted in any given time period is limited by the energy available during the same time period. We argue that there exists a *maximum sustainable throughput* in wireless sensor networks under such energy constraints. Our investigation towards the operational lifetimes of sensor networks also leads to significant results with respect to the maximum sustainable throughput. The following theorem establishes our key observation with respect to the maximum sustainable throughput.

Theorem II-A 2.

(1) The maximum sustainable throughput of a wireless sensor network with renewable energy sources is limited by n in the order of $b_\epsilon/\sqrt{n}$ for fixed network sizes and b_ϵ/n for fixed node densities.

(2) The energy-constrained capacity scales worse than interference-constrained capacity if the size of the network grows with n in the order greater than $\log n$.

The above results imply that, if the growth of the network size is not exceedingly slow compared to the growth of n, and when the number of nodes in the network is sufficiently large, the fundamental performance limits with respect to network capacity are dominated by the energy-constrained capacity, rather than interference-constrained information-theoretic capacity. The detailed proof of this theorem is postponed to Sec. II-F.

[1] Detailed justifications for such a definition of b_ϵ are postponed to Lemma II-D 1.

B. Problem Setup

We begin our journey towards proving the correctness of our key observations and claims previously stated. Fig. 1 illustrates the setup of the problem. Without loss of generality, consider a wireless sensor network with n nodes uniformly deployed within a square area of size A as shown in Fig. 1. Each node has constrained energy sources. Optimal power schedules are assumed, achievable by optimal power management strategies. The objective is to find asymptotic impact of constrained energy resources on the fundamental performance limits of the network, such as the operational lifetime and data transmission capacity.

We consider the typical scenario in which all nodes behave as sources and the data sink is the destination of all transmissions. In the case that the source can not reach the destination directly, intermediate nodes will act as relays to forward messages to the destination via multiple hops.

Fig. 1. We study a wireless sensor network with n nodes in a square area. The goal is to examine the asymptotic impact of constrained energy resources on the fundamental performance limits of the sensor network. The size of the square area is A.

As shown in Fig. 1, we consider nodes near y-axis (or x-axis) cuts close to the data sink. [2] Such a cut-based technique has been used in [10] and [11]. We argue that the analysis of the energy cost and thus the lifetime of the nodes near the cut can provide adequate insights into the energy cost and performance of the entire sensor network.

C. Assumptions and Definitions

Before we progress to a position to analyze the energy cost of data transmissions and the operational lifetime of the network, we need to clarify our basic assumptions and establish a few terms by definitions.

We first define the radio model used in this paper. In this paper, we adopt the first order radio model used in [2], [14] and many other literatures. In this model, the following energy parameters are included: transmit (α_{11}), receive (α_{12}), and transmit amplify (α_2). Without loss of generality, all respective energy costs are for one bit. We do not include the energy costs of sensing. The reason is that energy costs for sensing depends heavily on the specific application. Nevertheless, such energy costs can be easily integrated into our solution once the sensing model is defined. Based on such a radio model, the energy costs for transmitting the signal across the distance of r is

[2]The y-axis and x-axis cuts will be formally defined in Sec. II-C.

$$E_r = \alpha_{11} + \alpha_2 r^\alpha + \alpha_{12} = \alpha_1 + \alpha_2 r^\alpha \quad (1)$$

where $\alpha_1 = \alpha_{11} + \alpha_{12}$. When a source sends a message to a destination whose distance from source is d, it can use intermediate nodes to relay the message. Under the first order radio model, the optimal distance between relay nodes is the *characteristic distance* denoted by d_m [14]. d_m is defined as

$$d_m = \sqrt[\alpha]{\frac{\alpha_1}{\alpha_2(\alpha - 1)}} \quad (2)$$

d_m is independent of the source-destination distance d. Theorem 2 of [14] proves that d_m is the optimal hop distance for any d and the optimal number of hops taken, K, is given by either $K = \lfloor \frac{d}{d_m} \rfloor$ or $K = \lceil \frac{d}{d_m} \rceil$. Without loss of generality, we assume $d \gg d_m$ in our paper to facilitate discussions. Hence $K = \frac{d}{d_m}$.

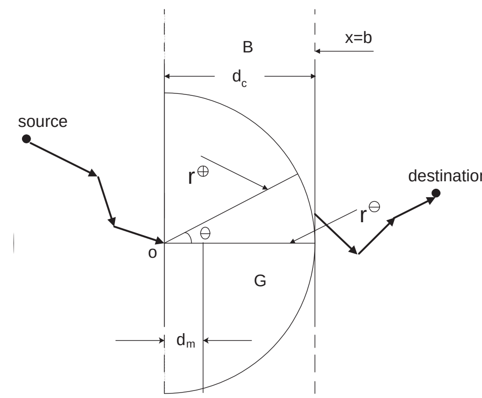

Fig. 2. A y-axis cut is placed at position $x = b$. $B = [b - d_c, b] \times [0, \sqrt{A}]$. Any message from source to destination must be relayed by one or more nodes in B.

We then formally define the concept of *cut* and *failed cut*. A y-axis cut at position x is a line segment parallel to the y-axis whose x-axis position equals x. The y-position of the line segment starts from 0 and ends at $\sqrt{A}$. When the nodes near the cut fail, network traffic can not cross such a y-axis cut. Such a y-axis cut is referred to as a *failed* y-axis cut. An x-axis cut at position y and a failed x-axis cut can be similarly defined.

We now consider the energy cost of nodes near a y-axis cut. As illustrated in Fig. 2, a y-axis cut is placed at position $x = b$. We are interested in the energy costs of the nodes in a set R, where R is defined as

$$R = \{v | (v_x, v_y) \in [b - d_c, b] \times [0, \sqrt{A}]\} \quad (3)$$

where d_c is the maximum transmission range used by the nodes in the network. In the case that each node uses different maximum transmission ranges, since d_c must apply to all the cuts, d_c should be interpreted as the average of the maximum transmission ranges used by all nodes. Informally, R represents the nodes within the rectangular area immediately left of b, whose width is d_c. Denote the area occupied by R as B. We have $B = [b - d_c, b] \times [0, \sqrt{A}]$.

Definition II-C 1. The *relay position set* G is defined as the set of possible positions of relay nodes.

Without loss of generality, assume that the transmission is from left to right. Then G is the right half of the circle centered at the sender. We assume that each node in the network needs to send 1 bit of information in each time slot. Since d_c is the maximum transmission range, any source must rely on one or more nodes in R to relay the messages in order to cross cut b. Let h_b be the number of hops a 1-bit message needs to take in B in order to cross cut b. Let r_j be the distance traversed in the jth hop, where $1 \leq j \leq h_b$. Let o_j and z_j be the sender and receiver of the jth hop, where obviously $z_{j-1} = o_j$ for $2 \leq j \leq h_b$. Let c_b be the energy spent by $o_j, 1 \leq j \leq h_b$. c_b is in fact the total energy spent by $v \in R$ in order to forward such 1-bit message to cross cut b. We then have

$$c_b = \sum_{j=1}^{h_b} E_{r_j} = \sum_{j=1}^{h_b} (\alpha_1 + \alpha_2 r_j^{\alpha}) \tag{4}$$

r_j are in fact iid random variables. This claim is based on the observation that *sub-paths of shortest paths are shortest paths*. Further, the setup of cut b and area B is artificial. The optimal schedule of each hop does not depend on the result of the previous hop. The only limitation is that the maximum transmission range can not exceed d_c. Another way to interpret this is that each hop actually belongs to multiple cuts.

Since r_j are iid random variables, E_{r_j} are also iid random variables. Therefore, whenever appropriate, we will omit the subscript j and use r and E_r respectively in later discussions. For instance, we can write

$$E_r = \alpha_1 + \alpha_2 r^{\alpha} \tag{5}$$

For practical applications, $d_m \ll d_c$. In addition, we are able to show that[3], the probability of $r \gg d_m$ is very low. Therefore we have the equality[4]

$$d_c = \sum_{j=1}^{h_b} r_j \cos(\theta_j) \tag{6}$$

where θ_j is the angle between $\vec{\mathbf{r}_j}$ and the x-axis. We are interested in the expectation of c_b under optimal power management. From Eq. (4), we have

$$E[c_b] = E[h_b]E[E_r] = \frac{d_c}{E[r\cos(\theta)]} E[E_r] \tag{7}$$

Lemma II-C 1. c_b satisfies

$$\frac{d_c(\alpha_1 + \alpha_2 d_m^{\alpha})}{d_m} \leq c_b \leq \frac{d_c(\alpha_1 + \alpha_2 d_c^{\alpha})}{d_m} \tag{8}$$

$$E[c_b] < \frac{d_c(\alpha_1 + \alpha_2 \omega_{\alpha}(\lambda, d_c))}{d_m(1 - e^{-\lambda \frac{\pi d_c^2}{2}})} \leq \frac{d_c(\alpha_1 + \alpha_2 d_c^{\alpha})}{d_m(1 - e^{-\lambda \frac{\pi d_c^2}{2}})} \tag{9}$$

where $\lambda = \frac{n}{A}$ and $\omega_{\alpha}(\lambda, d_c) < d_c^{\alpha}$. Asymptotically, $\omega_{\alpha}(\lambda, d_c)$ satisfies a) $\lim_{n\to\infty} \omega_{\alpha}(\lambda, d_c) = (1+\kappa)d_m^{\alpha}$, $\kappa > 0$; or b) $\omega_{\alpha}(\lambda, d_c)$ grows with n.

Proof: The proof of Eq. (8) is trivial. As proved in [14], the relay path with the least energy cost is the straight line parallel to the x-axis (Fig. 2). In addition, the distances traversed by each hop must equal to the characteristic distance d_m defined in Eq. (1) in order to achieve the minimum energy cost. Consequently, the minimum energy cost to relay a 1-bit message across the cut shown in Fig. 2 is $\frac{d_c(\alpha_1+\alpha_2 d_m^{\alpha})}{d_m}$. This minimum value can be achieved iff network nodes occupy the positions whose distances from the point o are multiples of d_m, which is not the case in general. The proof of Eq. (9) is much more involved. Due to lack of space, the detailed proofs are not included. However, the asymptotic results of this paper actually do not depend on the details of the closed form of ω_{α}. In fact, the qualitative claims pertaining to ω_{α} can be explained intuitively. Since d_m is the optimal distance [14], we can use $d_m(1 - e^{-\lambda \frac{\pi d_c^2}{2}})$ as the first order approximation of $E[r\cos(\theta)]$. Since $E_r = \alpha_1 + \alpha_2 r^{\alpha}$, the expectation $E[E_r]$ must have the form of $\alpha_1 + \alpha_2\omega_{\alpha}$, where $\omega_{\alpha} = E[r^{\alpha}]$. Obviously, $\omega_{\alpha} \leq d_c^{\alpha}$. Because d_m is the optimal distance, if the network node density increases with larger n (λ increases), the optimal distance d_m will more likely be chosen. In such cases, $\lim_{n\to\infty} \omega_{\alpha}(\lambda, d_c) = (1+\kappa)d_m^{\alpha}$. For $\alpha = 2$, it can be proved that $\kappa = \frac{1}{2\pi}$. if the network node density decreases with larger n (λ decreases), ω_{α} increases since $d_c \gg d_m$[5] However, ω_{α} never exceeds d_c^{α}. □

Fig. 3 shows the comparison of our theoretical bounds and simulation results. The energy parameters used in simulation and theoretical bounds are $\alpha_1 = 50$nJ/bit, and $\alpha_2 = 0.1$nJ/bit/m^2. The lower bound shown is calculated based on the assumption that the node at the optimal distance d_m along the straight line parallel to the x-axis will always be chosen as the relay node. As shown in this figure, the upper bound derived above is reasonable tight.

D. Operational Lifetime of Sensor Networks

In this paper, we are interested in the operational lifetime t_c of the sensor network. The nature of t_c can be explained as follows. With the progress of time, some of the network nodes may fail after the depletion of their energy resources. Even though the distributions of the failed nodes are random, there exist certain probabilities that some nodes that are physical proximate, such as the nodes in the region B near a cut (Fig. 2), may fail faster than some of the other nodes in the network. If this event happens, all the network transmissions that across that cut will break. When the number of the failed cuts in certain critical region (as will be analyzed in Fig. 4) reach certain threshold, more than $100(1-\epsilon^2)\%$ data transmissions fail and the network reach its operational life. Our study of operational lifetime t_c explores the aforementioned

[3]Detailed proofs are not included due to space constraints.

[4]In general, the inequality $d_c \leq \sum_{j=1}^{h_b} r_j \cos(\theta_j) \leq 2d_c$ rather than the equality in Eq. (6) holds. For simplicity of presentation, we introduce the equality here. The asymptotic results of this paper remain the same if the inequality is applied.

[5]If $d_c \ll d_m$, ω_{α} will approach a constant > 0, because $d_m > 0$ is the optimal distance.

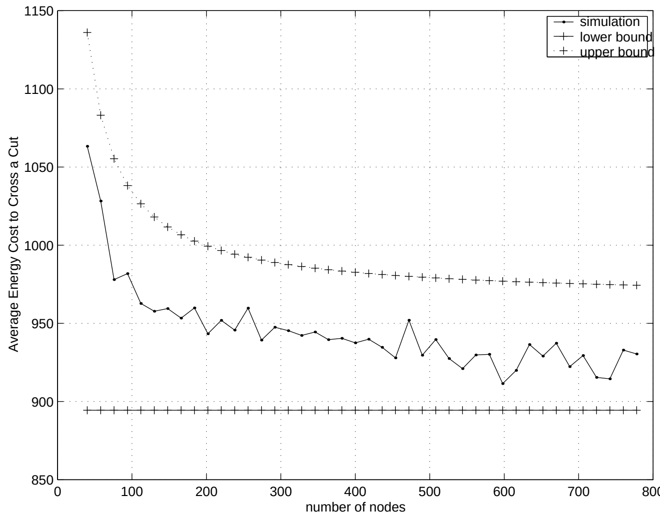

Fig. 3. Comparison between simulation results and theoretical lower and upper bounds, where $d_c = 200$m, and $A = 4 \times 10^6 \mathrm{m}^2$. For convenience of illustration, the number of nodes shown is $\sqrt{n}$ rather than n. The unit of y-axis is nJ. $\alpha = 2$.

characteristics of the network. In general, the operational lifetime can be reached long before the last node in the network fails.

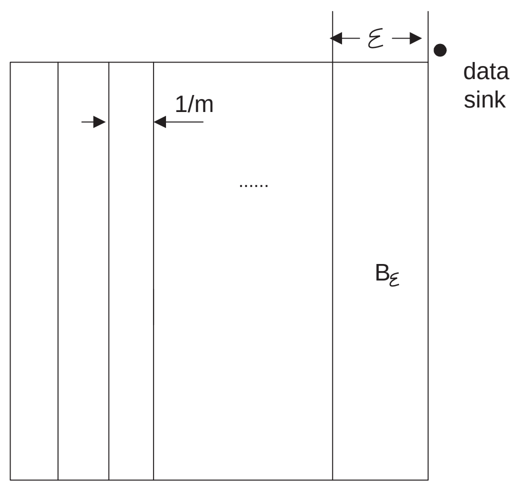

Fig. 4. There are m y-axis cuts, whose positions are $\frac{i}{m}\sqrt{A}, 1 \le i \le m$. For the convenience of illustration, the figure shown above has been normalized to the size of 1. When considering y-axis only, $100\epsilon\%$ transmissions will break iff there is at least one failed y-axis cut whose position $x \ge (1-\epsilon)\sqrt{A}$.

As assumed in Sec. II-C, each node in the network needs to send 1 bit of information in each time slot. As shown in Fig. 4, we assume that there are m y-axis cuts, whose positions are $\frac{i}{m}\sqrt{A}, 1 \le i \le m$. The value of m will be determined later in the proof of Lemma II-D 1. For the cut at the position $\frac{i}{m}\sqrt{A}$, the amount of traffic needs to be relayed is $\frac{i}{m}n$. Therefore, the total energy costs in t time slots for nodes $v \in R$ near such a cut is

$$c_{t,i} = \sum_{k=1}^{t} \frac{i}{m} n c_{b,k,i} \tag{10}$$

where $c_{b,k,i}$ is the average energy cost of, in the kth time slot, to forward a 1-bit message across cut i. During the development of the upper bound of c_b, we consider the probability of occupying, the probability that a position in region B (Fig. 2) is occupied by a node, based only on the initial deployment. In practice, the probability of occupying is affected by the energy cost as well. A node will fail after depleting its energy resources. Because of the failed nodes, the effective number of the nodes in the network decreases over time. Therefore, to model $c_{b,k,i}$, the energy cost in the network after the number of nodes decreases, the value of λ needs adjustment accordingly. λ should equal to $\frac{n}{A}\tau_{k,i}$, where $0 < \tau_{k,i} \le 1$. However, because the initial distribution of the nodes in the network is random and the random nature of the network traffic, the distributions of the failed nodes are also random. Consequently, the results from Lemma II-C 1 still hold for $c_{b,k,i}$. In fact, the inequality in Eq. (8) and the qualitative claims pertaining to the asymptotic behavior ω_α hold even for uneven distributions of network nodes. In practice, we are interested in the values of n and A smaller than some finite values $n_{\max}$ and $A_{\max}$. Because a smaller λ leads to a larger c_b, there exists a value $\tau_{\min}$, whose corresponding $c_{b,\max}$ satisfies

$$E[c_{b,\max}(\frac{n}{A}\tau_{\min}, d_c)] \ge \max_{i,k,n<n_{\max},A<A_{\max}} E[c_{b,k,i}] \tag{11}$$

In addition, the value of $c_{b,\max}$ is independent from t_c. $c_{b,\max}$ depends on the node distribution characterized by $\lambda = \frac{n}{A}\tau_{\min}$. Therefore, we have

$$E[c_{t_c,i}] \le \frac{i}{m} n t_c E[c_{b,\max}] \tag{12}$$

In the remainder of the paper, all references to c_b are actually the references to $c_{b,\max}$. For simplicity, we drop the subscript $\max$ and use c_b hereafter.

Lemma II-D 1. The operational lifetime t_c satisfies $\frac{2b_\epsilon e_o}{\sqrt{nA}[\alpha_1+\alpha_2\omega_\alpha(\lambda,d_c)]} d_m(1-e^{-\lambda\frac{\pi d_c^2}{2}}) < t_c \le \frac{2\epsilon b_\epsilon e_o d_m}{\sqrt{n}d_c(\alpha_1+\alpha_2 d_m^\alpha)}$.

Proof: We consider the y-axis cuts first. Because the data sink is the destination of all the transmissions, $100\epsilon\%$ transmissions will break iff there is at least one failed y-axis cut whose position $x \ge (1-\epsilon)\sqrt{A}$. As shown in Fig. 4, we define the region covering these cuts as B_ϵ.

We develop the lower bound first. Let $n_f(t)$ be the number of failed cuts in region B_ϵ by time t, we then have

$$n_f(t) = \sum_{i=1}^{m\epsilon} I_{t,m-i+1} \tag{13}$$

where $I_{t,m-i+1}$ is the indicator variable defined as

$$I_{t,j} = \begin{cases} 1, & \text{cut at the position } \frac{j}{m}\sqrt{A} \text{ has failed by time } t. \\ 0, & \text{cut at the position } \frac{j}{m}\sqrt{A} \text{ is active by time } t. \end{cases} \tag{14}$$

$E[n_f(t)] = \sum_{i=1}^{m\epsilon} E[I_{t,m-i+1}] = \sum_{i=1}^{m\epsilon} P[I_{t,m-i+1} = 1]$, where, based on Markov inequality, $P[I_{t,m-i+1} = 1] = P[c_{t,m-i+1} \ge nd_{cr}e_o] \le \frac{E[c_{t,m-i+1}]}{nd_{cr}e_o}$, where e_o is the initial energy available at each node, and $d_{cr} = \frac{d_c}{\sqrt{A}}$. Hence t_c, the expected time required for $100 \times (1-\epsilon)\%$ network transmissions to fail due to failed y-axis cuts satisfies the following:

$$\begin{aligned} 1 \leq E[n_f(t)] &\leq \sum_{i=1}^{m\epsilon} \frac{E[c_{t,m-i+1}]}{nd_{cr}e_o} \\ &\leq \frac{t_c E[c_b]}{d_{cr}e_o} \sum_{i=1}^{m\epsilon} \frac{m-i+1}{m} \\ &= \frac{t_c E[c_b]}{d_{cr}e_o} m\epsilon[(1+\frac{1}{m}) - \frac{1}{2}\epsilon - \frac{1}{2m}] \end{aligned}$$

Since there are n nodes in the network, we then have $m = \sqrt{n}$ [6]. Therefore, as $n \to \infty$, t_c satisfies

$$1 \leq \frac{t_c E[c_b]}{d_{cr}e_o} \sqrt{n}\epsilon(1 - \frac{1}{2}\epsilon)$$

Let $b_\epsilon = \frac{1}{\epsilon(2-\epsilon)}$, we then have

$$\begin{aligned} \lim_{n\to\infty} t_c &\geq \frac{2b_\epsilon d_{cr} e_o}{\sqrt{n}E[c_b]} \\ &> \frac{2b_\epsilon e_o}{\sqrt{nA}[\alpha_1 + \alpha_2\omega_\alpha(\lambda, d_c)]} d_m(1 - e^{-\lambda\frac{\pi d_c^2}{2}}) \end{aligned}$$

Using the second inequality of Eq. (8), we can follow the steps similar to the above derivation process, but without the need to reference Eq. (11) and Eq. (12), to develop a relatively loose lower bound of t_c as follows:

$$\lim_{n\to\infty} t_c > \frac{2b_\epsilon e_o}{\sqrt{nA}(\alpha_1 + \alpha_2 d_c^\alpha)} d_m$$

To develop the upper bound, we use the condition that the total energy cost of all the cuts in the region B_ϵ can not exceed $n\epsilon e_o$. We then have

$$\begin{aligned} n\epsilon e_o &\geq \sum_{i=1}^{m\epsilon} c_{t,m-i+1} \\ &= \sum_{i=1}^{m\epsilon}\sum_{k=1}^{t_c} \frac{m-i+1}{m} nc_{b,k,i} \end{aligned}$$

Applying Eq. (8) of Lemma II-C 1, we have

$$\begin{aligned} n\epsilon e_o &\geq \sum_{i=1}^{m\epsilon}\sum_{k=1}^{t_c} \frac{m-i+1}{m} n \frac{d_c(\alpha_1 + \alpha_2 d_m^\alpha)}{d_m} \\ &= nt_c \frac{d_c(\alpha_1 + \alpha_2 d_m^\alpha)}{d_m} \sum_{i=1}^{m\epsilon} \frac{m-i+1}{m} \end{aligned}$$

Applying $m = \sqrt{n}$ as before, we have

$$\lim_{n\to\infty} n\epsilon e_o \geq nt_c \frac{d_c(\alpha_1 + \alpha_2 d_m^\alpha)}{d_m} \sqrt{n}\epsilon(1 - \frac{1}{2}\epsilon)$$

Therefore, the upper bound of t_c is

$$\lim_{n\to\infty} t_c \leq \frac{2\epsilon b_\epsilon e_o d_m}{\sqrt{n} d_c(\alpha_1 + \alpha_2 d_m^\alpha)} \tag{15}$$

The upper bound just derived is a loose bound. To achieve the equality in Eq. (15), two conditions must be met. First, all the energy resources on all the nodes in the region B_ϵ must be depleted at the time t_c. Second, the minimum energy cost $\frac{d_c(\alpha_1+\alpha_2 d_m^\alpha)}{d_m}$ must be achieved for all data transmissions throughout the lifetime of the network, which is not possible at the later stage of network life. Nevertheless, the upper bound in Eq. (15) offers additional proof that the energy-constrained capacity scales worse than the interference-constrained capacity. For instance, for fixed λ, the upper bound of t_c scales as $\frac{1}{\sqrt{n \log n}}$ because d_c must grow with n in the order of $\Theta(\sqrt{\frac{\log n}{\pi n}})$ in order to keep the network connected [16]. Note that in Gupta and Kumar [16], it is assumed that the network occupies unit area. Compare $\frac{1}{\sqrt{n \log n}}$, the upper bound of t_c for fixed λ with the interference-constrained capacity $\Theta(\frac{1}{\sqrt{n \log n}})$ [17], [10], [22], we prove that, since the upper bound of t_c is not attainable, the energy-constrained capacity scales worse.

In the above derivations, we only consider the broken network transmissions due to failed y-axis cuts. In fact, when the expected number of failed y-axis cuts reaches 1, the expected number of failed x-axis cuts in B_ϵ also reaches 1. Therefore, at time t_c, there are at least $100 \times (1-\epsilon^2)\%$ network transmissions broken.

Let t_{low} represent the lower bound of the network lifetime, we then have

$$t_{low} = \frac{2b_\epsilon e_o}{\sqrt{nA}[\alpha_1 + \alpha_2\omega_\alpha(\lambda, d_c)]} d_m(1 - e^{-\lambda\frac{\pi d_c^2}{2}}) \tag{16}$$

□

E. Proof of Theorem II-A 1

We are now in the position to prove Theorem II-A 1.

Proof:

1) Fixed A, variable n: In such a scenario, the node density varies while the deployment area of the sensor network remains the same. As shown in the proof of Lemma II-C 1, $\lim_{n\to\infty} \omega_\alpha(\lambda, d_c)$ =constant. Using Lemma II-D 1, we have $t_{low} \propto \frac{2b_\epsilon e_o}{\sqrt{nA}}$, when n is large. Therefore, t_{low} will decrease in the order of $n^{-\frac{1}{2}}$.

2) Fixed λ, variable n: In such a scenario, the size of the network varies while the node density remains the same. When λ is fixed, we are able to prove that[7], $\lim_{n\to\infty} \omega_\alpha(\lambda, d_c)$ =constant for both fixed and variable d_c by using the result that d_c needs to grow with n in the order of $\Theta(\sqrt{\frac{\log n}{\pi n}})$ in order to keep the network connected[16][8]. Thus $t_{low} \propto \frac{1}{\sqrt{nA}} = \frac{\sqrt{\lambda}}{n}$. That is, t_{low} decreases in the order of $\frac{1}{n}$ when the coverage of the sensor network grows while node density remains the same.

[6] A more precise estimate of m is $m \propto \sqrt{n}$. For simplicity of discussions, we let $m = \sqrt{n}$, since it will not affect our discussions of the asymptotic behavior of t_c.

[7] Detailed proofs are not included due to space constraints.

[8] In Gupta and Kumar [16], it is assumed that the network occupies unit area.

3) Comparison with the Average Node Lifetime: Let t_o represent the average lifetime of an individual node. Then $t_o \propto \frac{nd_{cr}e_o}{nE[c_b]/2}$. Compare this with the result from Lemma II-D 1, we conclude that the operational lifetime of a wireless sensor network is shorter than the average lifetime of an individual node by a factor of $b_\epsilon/\sqrt{n}$ for fixed network sizes and b_ϵ/n for fixed node densities. Since ϵ can not be too small, b_ϵ can not be too large. Therefore, the operational lifetime is much smaller than the average lifetime of an individual node when the number of nodes in the network is large. □

Intuitively, for fixed network sizes, when there are more nodes generating traffic, there are more nodes available to relay the traffic across the cut because the node density will grow with n. Under optimal power management, when the number of nodes is sufficiently large, the characteristic distance will always be chosen. Thus the network lifetime depends mainly on the number of cuts vulnerable. We can then conclude that adding more nodes in the initial deployment does not add redundancy, since each new node needs to generate traffic and relay traffic for other nodes. Note that as n grows, the absolute number of transmissions remaining after t_c is larger, but it is much easier for a certain percentage of transmissions to fail when n is larger.

When the network size needs to grow with the number of nodes, the node density remains the same. Therefore, there will be relatively fewer nodes available to relay the growing traffic across the cut. Such disparity will grow in the order of $\frac{1}{\sqrt{n}}$. Together with the $\frac{1}{\sqrt{n}}$ factor introduced by the number of vulnerable cuts, the lifetime decreases with n in the order of $\frac{1}{n}$.

The above discussion leads to the significance of the characteristic distance. If $\alpha_1 = 0$ (in which case the characteristic distance is not significant), more nodes will always lead to less energy costs and a longer lifetime. However, when α_1 is significant, the relay energy cost will remain the same after n reach a certain threshold.

The third result of Theorem II-A 1 confirms the validity of the definition of the operational lifetime proposed in this paper. It shows that the network has ceased to function long before the last node fails from energy depletion.

F. Proof of Theorem II-A 2

Using Lemma II-D 1, we are now ready to prove Theorem II-A 2.

Proof:

1) Scalability of Maximum Sustainable Throughput: For sensor networks depending on renewable energy such as solar energy, the maximum amount of data that can be transmitted in any given time period is limited by the energy available during the same time period. Let w denote the *maximum sustainable throughput, i.e.*, the maximum number of bits can be injected into the network by each node without causing network failure as a result of energy depletion. Based on Eq. (16), w can be found by solving following equation:

$$t = \frac{2b_\epsilon e_s t d_m (1 - e^{-\lambda \frac{\pi d_c^2}{2}})}{w\sqrt{nA}(\alpha_1 + \alpha_2 \omega_\alpha(\lambda, d_c))}$$

$$w = \frac{2b_\epsilon e_s d_m (1 - e^{-\lambda \frac{\pi d_c^2}{2}})}{\sqrt{nA}(\alpha_1 + \alpha_2 \omega_\alpha(\lambda, d_c))} \quad (17)$$

where e_s is the power renewal rate. Therefore, $w \propto \frac{b_\epsilon}{\sqrt{n}}$ for fixed network sizes and $w \propto \frac{b_\epsilon}{n}$ for fixed node densities.

2) Comparison with Interference-Constrained Capacity: We compare the above result with the capacity predicted based on the interference among concurrent transmissions. Interference-constrained capacity per node scales as $\Theta(\frac{1}{\sqrt{n \log n}})$ [17], [10], [22]. When λ is fixed, it is obvious that the energy-constrained capacity scales much worse than interference-constrained capacity. In fact, $\omega_\alpha(\lambda, d_c)$ either approaches a constant or grows with n, and $(1 - e^{-\lambda \frac{\pi d_c^2}{2}})$ either approaches zero or a constant no greater than 1. Therefore, as long as A grows with n in the order greater than $\log n$, the maximum sustainable throughput w scales worse than interference-constrained capacity. In this case, when the number of nodes is sufficiently high, the energy-constrained network capacity dominates. □

For a fixed A, since $\sqrt{\log n}$ grows very slowly with n, the scalability of the energy-constrained capacity and interference-constrained capacity with respect to n are comparable. Therefore, in the case that the power of the renewable energy source is constrained compared to the power consumption of the system (after the adjustments necessary for considering other variables and constants in Eq. (16)), if due to technology advances, the raw system transmission capacity of the sensors grows much faster than the system power efficiency, or the system power is increased in order to produce higher network throughput (as proposed in [22]), the energy-constrained capacity will dominate.

Because of constrained energy levels, the feasibility of deploying sensor networks has to be studied prior to its deployment. We can use Eq. (16) and Theorem II-A 1 and II-A 2 to calculate the expected operational lifetime of the network. In such cases, there usually exists minimum coverage requirements on node density (for fixed A) or area covered (for fixed λ). There may also exist minimum requirements on the network throughput. If the operational lifetime calculated based on such minimum coverage requirements can not cover the entire mission, a network replenishment strategy has to be devised to add additional nodes — thus more energy — into the network to ensure that the network will complete its mission successfully. Using Eq. (16) and constraints such as the deployment cost and the time window, a linear programming problem can be formulated to identify the optimal timing, location, size of node additions, and the schedule and amount of data transmissions, while minimizing the costs and risks involved.

III. RELATED WORK

In multi-hop wireless networks such as sensor networks, much efforts have been devoted to the problems of topology control, power schedule and optimal routing. For example, [6] studied topology control, [3], [5] studied energy-aware routing in wireless ad hoc networks, [15] studied minimum energy cost problems for broadcast and multicast, and [1],

[2] studied energy management in wireless sensor networks. Recently, [12] presented a competitive and efficient algorithm for routing of messages in energy-constrained ad-hoc network.

The paper most related to our work is written by Bhardwaj *et al.* [14]. In this paper, they studied the upper bound of lifetime of sensor networks with a single data source. However, the lifetime studied in this paper is the *active* lifetime, *i.e.*, the time at which the total energy consumed equals the the total energy in the network available at the start. As shown in our paper, such definition of upper bound yields very little relevance to the practical network. Since the network fails to function long before the last node in the network fails from energy depletion. In addition, [14] uses very simple models that fail to consider the stochastic behavior of relay nodes along the path between the source and data sink. Another related work is by Shakkottai *et al.* [13]. In this work, it is shown that the the necessary and sufficient conditions for the coverage and the connectivity of the random grid network are $p(n)r^2(n) \sim \frac{\log(n)}{n}$. The authors claimed that, using a node failure model as a function of time, the results in the paper can be used to answer the question pertaining to the maximum length of time over which one can expect the network to provide coverage and be connected with probability no smaller than $1-\epsilon$, where ϵ is a small value.

There have been many papers on the capacity of wireless ad hoc networks. These papers focus on the interference-constrained capacity of the network. The comparison of interference-constrained capacity and energy-constrained capacity has been presented in Sec. II-F. Gupta *et al.* [17] predicted that the total throughput of the network is $\Theta(\sqrt{n/\log n})$. Grossglauser *et al.* [18] proposed that mobility can increase the capacity of the wireless ad hoc networks. Li *et al.* [19] demonstrated that the scalability of mobile ad hoc networks depends on whether the network traffic can be localized. Peraki *et al.* [10] and Scaglione *et al.* [11] used the cut-based analysis. In both papers, the maximum data capacity that can cross any cut is studied to find the maximum network throughput based on max-flow/min-cut theorem of the flow network. In particular, Peraki *et al.* studied the maximum stable throughput problem in random networks with directional antennas. Barros *et al.* [20] studied the reachback capacity of sensor networks. Servetto [21] investigated the feasibility of large scale sensor networks. Xie *et al.* [22] studied the optimal strategies for information transmission and cooperation among the network nodes. One of the conclusions in [22] is that the transport capacity of the network is upper bounded by the total power for a positive absorption constant or path loss exponent greater than 3.

IV. CONCLUSION

In this paper, we studied the asymptotic behavior of operational lifetime and energy-constrained capacity of sensor networks. For sensor networks with renewable energy sources, our analysis shows that the maximum sustainable throughput in the network scales much worse than the capacity predicted based on interference among concurrent transmissions, if the growth of the physical network size is not exceedingly slow compared to the growth of n. Our results can be used to study the feasibility of deploying energy-constrained sensor networks and their replenishment strategies.

REFERENCES

[1] C. Schurgers, M. Srivastava. “Energy Efficient Routing in Wireless Sensor Networks,” in Proc. MILCOM 2001, pp. 357-361, October 2001.

[2] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. “Energy-Efficient Communication Protocols for Wireless Microsensor Networks,” in Proc. Hawaaian Int’l Conf. on Systems Science, January 2000.

[3] S. Singh, M. Woo, and C. Raghavendra. “Power-Aware Routing in Mobile Ad Hoc Networks,” in Proc. 4th ACM/IEEE Mobicom, 1998.

[4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. “A Survey on Sensor Networks,” IEEE Comm. Magazine, Vol. 40, No. 8, pp. 102-114, August 2002.

[5] V. Rodoplu and T. H. Meng. “Minimum Energy Mobile Wireless Networks,” IEEE Jour. Selected Areas Comm., vol. 17, no. 8, pp. 1333-1344, Aug. 1999.

[6] R. Wattenhofer, L. Li, V. Bahl and Y. M. Wang. “Distributed Topology Control for Power Efficient Operation in Multihop Wireless Ad Hoc Networks,” in Proc. IEEE INFOCOM, pages 1388-1397, April 2001.

[7] A. Leon-Garcia. “Probability and Random Processes for Electrical Engineering,” Addison-Wesley, 2nd Edition. Addison-Wesley, 1994.

[8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. “Introduction to Algorithms,” 2nd Ed., MIT Press, 2001.

[9] T. S. Rappaport. “Wireless communications: Principles and Practice,” Prentice Hall, 1996.

[10] C. Peraki, S. D. Servetto. “On the Maximum Stable Throughput Problem in Random Networks with Directional Antennas,” in Proc. of the 4th ACM MobiHoc, June 2003.

[11] A. Scaglione, S. D. Servetto. “On the Interdependence of Routing and Data Compression in Multi-Hop Sensor Networks,” in Proc. of the 8th ACM Mobicom, Atlanta, September 2002.

[12] K. Kar, M. Kodialam, T. V. Lakshman, and L. Tassiulas. “Routing for Network Capacity Maximization in Energy-constrained Ad-hoc Networks”, in Proc. of IEEE INFOCOM, pages 673-681, 2003.

[13] S. Shakkottai, R. Srikant and N. Shroff. “Unreliable Sensor Grids: Coverage, Connectivity and Diameter,” in Proc. of IEEE INFOCOM, pages 1073-1083, 2003.

[14] M. Bhardwaj, T. Garnett, and A. P. Chandrakasan. “Upper Bounds on the Lifetime of Sensor Networks,” in Proc. ICC 2001, June 2001.

[15] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides. “On the Construction of Energy-efficient Broadcast and Multicast Trees in Wireless Networks,” in IEEE INFOCOM 2000, pages 586-594, Tel Aviv, Israel, 2000.

[16] P. Gupta and P. R. Kumar. “Critical Power for Asymptotic Connectivity in Wireless Networks,” in Stochastic Analysis, Control, Optimization and Applications, Birkhauser, 1998.

[17] P. Gupta and P. R. Kumar, “The Capacity of Wireless Networks”, in IEEE Transactions on Information Theory, 46(2):388-404, March 2000.

[18] M. Grossglauser and D. Tse, “Mobility Increases the Capacity of Ad-hoc Wireless Networks.” in IEEE/ACM Transactions on Networking, vol. 10, no. 4, August, 2002, pp. 477-486.

[19] J. Li, C. Blake, D. Couto, H. Lee, R. Morris, “Capacity of Ad Hoc wireless networks,” in 7th ACM Mobicom, Rome, September 2001.

[20] J. Barros, S. D. Servetto. “Reachback Capacity with Non-Interfering Nodes,” in Proc. of the IEEE International Symposium on Information Theory (ISIT), Yokohama, Japan, June-July 2003.

[21] S. D. Servetto. “On the Feasibility of Large-Scale Wireless Sensor Networks,” in Proc. 40th Annual Allerton Conference on Communication, Control, and Computing, Urbana, IL, October 2002.

[22] Liang-Liang Xie and P. R. Kumar, “A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation,” Submitted to IEEE Transactions on Information Theory.

[23] C. Intanagonwiwat, R. Govindan and D. Estrin. “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks,” in Proc. 6th ACM Mobicom, Boston, Massachusetts, August 2000.