

Modeling Search Costs in Wireless Sensor Networks

Joon Ahn and Bhaskar Krishnamachari
Ming Hsieh Department of Electrical Engineering
Viterbi School of Engineering
University of Southern California, Los Angeles, CA 90089
Los Angeles, California, USA
joonahn@usc.edu, bkrishna@usc.edu

Abstract— We develop approximate closed-form expressions of expected minimum search energy costs for data-centric wireless sensor networks showing the search performance with respect to the network size N and the number of randomly placed copies of the target event r . We consider both unstructured sensor networks, which use blind sequential search for querying, and structured sensor networks, which use efficient hash-based querying. We also consider two kinds of deployments: a fixed transmit power (FTP) model and the geometric random graph (GRG) model. We find that the search cost of unstructured networks under the FTP deployment is proportional to N and inversely proportional to $(r + 1)$ regardless of the spatial dimension d in which nodes are deployed, while that of the GRG is proportional to $\frac{N(\log N)^{\frac{d}{\eta}}}{r+1}$ where η is the path-loss exponent. The search cost of structured networks under the FTP deployment is found to be proportional to $\sqrt[d]{N}/\sqrt[d]{r}$, while that of the GRG deployment is proportional to $\frac{\sqrt[d]{N(\log N)^{\eta-1}}}{\sqrt[d]{r}}$. In all cases, we also provide bounds on the coefficient of proportionality.

I. INTRODUCTION

There are two popular ways in which the wireless sensor network can be operated. One way is to be operated in a continuous data gathering mode, and the other way is to consider the wireless sensor network as a decentralized data-centric storage system. While the network in the continuous data-gathering mode is popular, mainly because it is easier to analyze and simpler to implement, continuous data gathering from all sensors is generally very inefficient if most of the sensed information is not essential, or if there are multiple sinks that may need different subsets of the sensed information at different times. In this case, however, the data-centric network might be more suitable for the energy efficiency. In such a data-centric approach, the sensed data can be stored either locally or at one or more remote locations within the network. Event information is obtained by sinks through queries that are issued on an on-demand basis.

When a sink knows where the nearest copy of the target event information is stored (e.g. using hash-based data centric storage techniques such as GHT [1], DIM [2], etc.), the search cost is the energy expenditure to send a query for the event to the target node and bring the information back through the shortest path. On the other hand, when a sink

has no clue where the target resides, it resorts to search for it blindly, which often is led to some sort of flooding. We call the network which adopts the former scheme the structured network and the latter scheme the unstructured network. While the structured network has smaller search energy cost at the cost of maintaining the location information of all events, the unstructured network doesn't need the overhead of maintenance.

In this work, we focus on deriving closed-form expressions for the expected minimum search energy cost considering structured and unstructured sensor networks deployed in a d -dimensional area. Since the energy is one of the most precious resources and searching is one of fundamental operations in the data-centric sensor network, it can be very useful to derive the closed-form expressions of such costs in the sense that they can be basis of other analytical modelings with tractable optimization. We have focused on analyzing search performance with respect to two key parameters - the size of the network and the number of copies of the event information.

In our analysis we consider two types of deployment models. In both cases we assume that the spatial density is fixed, so that the network deployment region grows proportionally to the number of nodes in the network. In the first model, called the fixed transmission power (FTP) model, the deployment is organized in such a way as to ensure that the network is connected despite each node operating with a fixed radio range (regardless of network size). In the second case, the nodes are deployed uniformly and independently at random; however, in this case, referred to as the geometric random graph (GRG) model, ensuring connectivity with high probability requires that the radio range be scaled with the network size to ensure that each node has a logarithmic number of neighbors on average [3], [4], [5].

There are several related works as follows. For the structured network, the minimum search cost is related with the shortest path between a sink and a source. For the unstructured network, we use the expanding ring search mechanism in which the sink sends a series of controlled floodings until it finds the event information. Chang and Liu [6] have found the way how to construct the series of controlled floodings in order to minimize the expected search cost given the distribution

of the event's location. And in our previous work [7], we had derived the approximate closed-form expression for the minimum search cost under 2-dimensional FTP node deployment (a special case of the more general results presented here). In [11], we have applied these search cost expressions to analyze the fundamental scaling behavior of data-centric sensor networks.

II. ASSUMPTIONS

The following are the key assumptions in our work.

- N nodes are deployed with a constant density in a d -dimensional ball \mathbb{B}^d space. The constant density implies that if the network size is increased, the deployment area grows proportionally. Initially, we consider the FTP model in which the radio range R is kept fixed, but the deployment is such that the network remains connected. Later, we consider the GRG model in section V, in which the radio range R is scaled proportionally with $\sqrt[d]{\log N}$ to ensure connectivity of the network with high probability.
- The distribution of events is assumed to be uniform in the deployment area.
- A total of r copies of an event are maintained with the uniform distribution in the network by creating $r - 1$ additional replicas when the event is first sensed.
- Each query is a one-shot query (i.e. requires a single response, not a continuous stream), and is satisfied by locating a single copy of the corresponding event.
- We assume that the links over which transmissions take place are lossless (e.g., using blacklisting) and present no interference due to concurrent transmissions (e.g., due to low traffic conditions or due to the use of a scheduled MAC protocol).
- For the FTP deployment model, the total energy cost for searching is assumed to be proportional to the total number of transmissions. For the GRG deployment model, we shall assume that the energy expenditure per transmission scales as R^η , where η is the path-loss exponent.
- We assume that the boundary effect is negligible.

III. STRUCTURED NETWORKS UNDER THE FTP MODEL

We first consider structured networks where nodes are deployed with constant node density ρ in the d -dimensional ball. We further assume that the network is sufficiently dense so that all nodes within a distance kR of the sink can be reached in k hops. The nodes in the network are all located within L hops of the sink. When modeling the search cost we assume that the sink is located in the center of the region. In our previous work [7], we have shown that relaxing this assumption does not provide big differences by simulation. Let $V_d(x)$ denote the volume of a d -ball of radius x , $N_d(h)$ the number of nodes at most h hop away from the sink. The volume of the ball is known to be expressed as follows:

$$V_d(x) = f(d) \cdot x^d \quad (1)$$

where $f(d) = \frac{2\pi^{d/2}}{d \cdot \Gamma(d/2)}$.

In this paper, $\Gamma(\cdot)$ is the Gamma function. Then, the number of nodes at most h hop away from the sink is given by,

$$N_d(h) = \rho f(d) \cdot (hR)^d = \tau(d) \cdot h^d \quad (2)$$

where

$$\tau(d) \doteq \rho f(d) R^d \quad (3)$$

which is the average number of neighbors of a node. Hence, the total number of nodes N can be expressed as follows:

$$N = N_d(L) = \tau(d) \cdot L^d \quad (4)$$

Now we recall that there are r number of copies of an event distributed uniformly randomly in the network. Let the random variable X_{min} denote the hop distance to the nearest copy of them from the querier. Its tail distribution is as follows:

$$\begin{aligned} P\{X_{min} > x\} &= \prod_{i=1}^r P\{i\text{-th copy is not in } x \text{ hop neighbors}\} \\ &= \left(1 - \frac{N_d(x)}{N}\right)^r = \left(1 - \frac{x^d}{L^d}\right)^r \end{aligned} \quad (5)$$

In the structured network, the search cost is related to a path of the lowest cost from a querier to the nearest node which has one of the copies. We assume the shortest path routing scheme so that the path would be their shortest path. Hence, the search cost is equal to the hop count from the querier to the nearest copy through the shortest path, which is denoted by X_{min} , plus the cost back to the querier. Hence, the expected search cost of the network deployed in d dimension is as follow:

$$C_{s,st}^{(d)} = 2 E[X_{min}] \quad (6)$$

Using the tail distribution given in Equation (5) and approximating summation to integration, we have

$$\begin{aligned} E[X_{min}] &= \sum_{x=0}^L P\{X_{min} > x\} \approx \int_0^L \left(1 - \frac{x^d}{L^d}\right)^r dx \\ &= \frac{L \cdot \Gamma(\frac{1}{d})}{d} \cdot \frac{\Gamma(r+1)}{\Gamma(r + \frac{1}{d} + 1)} \end{aligned} \quad (7)$$

Using Lemma 1 stated below and the equation $L = \frac{1}{\sqrt[d]{\tau(d)}}$ (from Equation (4)), we can calculate the lower and upper bounds of the search cost:

$$C_{s,st}^{(d)}(N, r) > l(d) \cdot \frac{\sqrt[d]{N}}{\sqrt[d]{r}} \quad (8)$$

$$C_{s,st}^{(d)}(N, r) < u(d) \cdot \frac{\sqrt[d]{N}}{\sqrt[d]{r}} \quad (9)$$

where

$$\begin{aligned} l(d) &= 2 \frac{\Gamma(\frac{1}{d}) \exp(\frac{1}{d})}{d \sqrt[d]{\tau(d)}} \left(\frac{d}{d+1}\right)^{\frac{3d+2}{2d}} \\ u(d) &= 2 \frac{\Gamma(\frac{1}{d}) \exp\left(\frac{1}{d} + \frac{12+d}{12(12+13d)}\right)}{d \sqrt[d]{e \tau(d)}} \end{aligned}$$

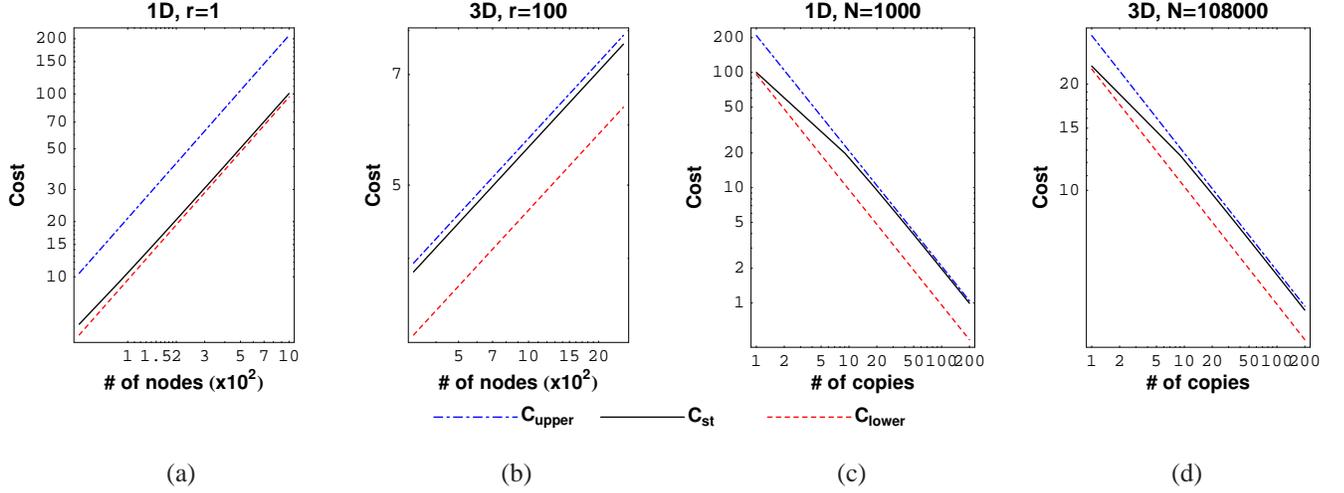


Fig. 1. Structured Average Search Cost

	$d = 1$	$d = 2$	$d = 3$
$l(d)$	0.0961058	0.410714	0.68273
$u(d)$	0.208858	0.577974	0.849538

TABLE I

THE COEFFICIENTS OF LOWER AND UPPER BOUNDS OF THE SEARCH COST FOR THE STRUCTURED NETWORK. THE NUMBER OF NEIGHBORS $\tau(d)$ OF A NODE IS SET TO 10.

Table I shows the numerically calculated values of $l(d)$ and $u(d)$ when the average number of one-hop neighbor $\tau(d)$ is 10 for 1, 2, and 3 dimensional deployments. As the table illustrates, the lower and upper bound are close and proportional to $\sqrt[d]{N}/\sqrt[d]{r}$ from the above double inequalities. Hence, we can approximate with good accuracy the search cost as follows:

$$C_{s,st}^{(d)} = \alpha_1 \cdot \frac{\sqrt[d]{N}}{\sqrt[d]{r}} \quad (10)$$

where $l(d) < \alpha_1 < u(d)$ (we can achieve the more accurate value of α using the curve fitting.)

Figure 1 verifies the accuracy of our model in various cases; (a) and (b) illustrate how the search cost varies as the number of nodes increases in 1D and 3D deployment, respectively. The number of copies of the event is 1 and 100 for (a) and (b), respectively. (c) and (d) show the search cost vs. the number of copies in 1D and 3D deployment, respectively. The number of nodes is 10^3 and 108×10^3 for (c) and (d), respectively. The bounds of the search cost are evaluated using Equation (8) and (9), and the search cost is evaluated numerically using Equation (6) with (7). All the four plots agrees that the search cost is almost proportional to the bounds in both relatively large and small networks, both 1D and 3D networks, and when r is small or large. As for 2-dimensional deployment, we have

investigated extensively in [7] and [8], which also agree with the above.

Lemma 1: For $r \geq 1$ and $1 \leq d \leq 12$, the following double inequality holds:

$$\left(\frac{d}{d+1}\right)^{\frac{2+3d}{2d}} \frac{\exp\left(\frac{1}{d}\right)}{\sqrt[d]{r}} < \frac{\Gamma(r+1)}{\Gamma\left(r+\frac{1}{d}+1\right)} < \frac{\exp\left(\frac{1}{d} + \frac{12+d}{12(12+13d)}\right)}{\sqrt[d]{e}} \cdot \frac{1}{\sqrt[d]{r}}$$

Proof: From Robbins 1955 [10], Stirling's approximation can be extended to the following double inequality:

$$\Gamma(r+1) > \sqrt{2\pi r} r^{r+\frac{1}{2}} e^{-r+\frac{1}{12r+1}} \quad (11)$$

$$\Gamma(r+1) < \sqrt{2\pi r} r^{r+\frac{1}{2}} e^{-r+\frac{1}{12r}} \quad (12)$$

Using Equation (12),

$$\Gamma\left(r+\frac{1}{d}+1\right) < \sqrt{2\pi} \left(r+\frac{1}{d}\right)^{r+\frac{1}{d}+\frac{1}{2}} e^{-r-\frac{1}{d}+\frac{d}{12rd+12}}$$

From the above equation and (11),

$$\begin{aligned} \frac{\Gamma(r+1)}{\Gamma\left(r+\frac{1}{d}+1\right)} &> \frac{1}{\sqrt[d]{r}} \left(\frac{rd}{rd+1}\right)^{\frac{2rd+d+2}{2d}} e^{\left(\frac{1}{d} + \frac{12-d}{12(12r+1)(rd+1)}\right)} \\ &\geq \frac{1}{\sqrt[d]{r}} \left(\frac{d}{d+1}\right)^{\frac{3d+2}{2d}} e^{\left(\frac{1}{d} + \frac{12-d}{12(12r+1)(rd+1)}\right)} \\ &\geq \frac{1}{\sqrt[d]{r}} \left(\frac{d}{d+1}\right)^{\frac{3d+2}{2d}} e^{\frac{1}{d}} \end{aligned}$$

Note that the second inequality holds since $\left(\frac{rd}{rd+1}\right)^{\frac{2rd+d+2}{2d}}$ is increasing with respect to r for $r \geq 1$ so that it has its minimum value at $r = 1$. And the third inequality holds since $\frac{12-d}{12(12r+1)(rd+1)} \geq 0$ for $r \geq 1$ and $d \leq 12$.

In the other hand, using the Robbins' double inequality in the other way around produces the following:

$$\frac{\Gamma(r+1)}{\Gamma\left(r+\frac{1}{d}+1\right)} < \frac{1}{\sqrt[d]{r}} \left(\frac{rd}{rd+1}\right)^{\frac{2rd+d+2}{2d}} e^{\left(\frac{1}{d} + \frac{d+12}{12r(12rd+d+12)}\right)} \quad (13)$$

Let $p(r) = \left(\frac{rd}{rd+1}\right)^{\frac{2rd+d+2}{2d}}$ and $q(r) = e^{\left(\frac{1}{d} + \frac{d+12}{12r(12rd+d+12)}\right)}$. Then, let's calculate the supremum of each of them.

$$\begin{aligned} \sup_{r \geq 1} p(r) &= \lim_{r \rightarrow \infty} \left(\left(1 - \frac{1}{rd+1}\right)^{rd+1} \right)^{\frac{1}{d}} \left(\frac{rd}{rd+1}\right)^{-\frac{1}{d}} \\ & (\because p(r) \text{ is increasing w.r.t } r \text{ for } r \geq 1) \\ &= \lim_{t \rightarrow \infty} \left(\left(1 - \frac{1}{t}\right)^t \right)^{\frac{1}{d}} \cdot \lim_{r \rightarrow \infty} \left(\frac{rd}{rd+1}\right)^{-\frac{1}{d}} \\ & (\because \text{substituting } t \doteq rd+1) \\ &= e^{-\frac{1}{d}} \end{aligned} \quad (14)$$

$$\begin{aligned} \sup_{r \geq 1} q(r) &= \exp\left(\frac{1}{d} + \frac{d+12}{12r(12rd+d+12)}\right) \Big|_{r=1} \\ & (\because \text{the exponent is decreasing w.r.t } r \text{ for } r \geq 1) \\ &= \exp\left(\frac{1}{d} + \frac{d+12}{12(13d+12)}\right) \end{aligned}$$

Hence, the RHS of inequality (13) can be further upper-bounded using the above supremums resulting in the following:

$$\frac{\Gamma(r+1)}{\Gamma(r+\frac{1}{d}+1)} < \frac{\exp\left(\frac{1}{d} + \frac{12+d}{12(12+13d)}\right)}{\sqrt[d]{e}} \cdot \frac{1}{\sqrt[d]{r}}$$

■

IV. UNSTRUCTURED NETWORKS UNDER THE FTP MODEL

We now consider unstructured networks. The search cost consists of the cost to locate the nearest copy and the cost to bring the data back to the querier using the shortest path. Since the latter cost is much smaller than the former, we here ignore the latter cost, which is equal to $E[X_{min}]$ given in (7). Hence, we derive the search cost expression using the optimal expanding ring-based flooding query [6], [7]. We consider the same d -Ball as a network deployment space as in section III. We first consider the lower bound of the optimal expected search cost. Suppose a querying node happens to know the hop distance X_{min} to the nearest copy of the desired event before disseminating queries. Then, the flooding cost up to X_{min} hops away is certainly the lower bound. The distribution of X_{min} is given in section III. Under our assumption the expression for the flooding cost up to h hops away is given by,

$$C_f^{(d)}(h) = 1 + N_d(h-1) \quad (15)$$

where $N_d(h)$ is the number of nodes up to h hops away given in section III.

The lower bound of the expected search cost is given by,

$$\begin{aligned} C_{s,lower}^{(d)} &= E[C_f^{(d)}(X_{min})] \\ &\approx \tau(d) \cdot E[X_{min}^d] \end{aligned} \quad (16)$$

In order to obtain the d -th moment of X_{min} , we make an approximation that X_{min} is continuous. The probability

density function of X_{min} is given by,

$$f_{X_{min}}(k) = \frac{rd}{L^d} k^{d-1} \left(1 - \frac{k^d}{L^d}\right)^{r-1} \quad (17)$$

Then, the d -th moment is given by,

$$\begin{aligned} E[X_{min}^d] &= \frac{rd}{L^d} \int_0^L k^{2d-1} \left(1 - \frac{k^d}{L^d}\right)^{r-1} dk \\ &= \frac{L^d}{r+1} \\ &= \frac{1}{\tau(d)} \cdot \frac{N}{r+1} \quad (\because (4)) \end{aligned} \quad (18)$$

Substituting Equation (18) into Equation (16) we have the following expression:

$$C_{s,lower}^{(d)} = \frac{N}{r+1} \quad (19)$$

Now, let us consider the upper bound of the optimal expected search cost. We note that any expected search cost with a specific search sequence vector (SSV) is the upper bound. We consider two search sequence strategies to achieve two upper bounds, from which we obtain a tighter upper bound in terms of order notation. Let us first consider the step-by-step expanding ring search (ERS) strategy where the SSV is $\{1, 2, 3, \dots, L\}$. The expected cost of this strategy is given by,

$$\begin{aligned} C_{s,subs}^{(d)} &= \sum_{k=1}^L C_f^{(d)}(k) P\{X_{min} > k-1\} \\ &= \sum_{k=1}^L (1 + \tau(d) \cdot (k-1)^d) \left(1 - \frac{(k-1)^d}{L^d}\right)^r \\ &\approx \int_0^L \tau(d) \cdot k^d \left(1 - \frac{k^d}{L^d}\right)^r dk \\ &= \tau(d) \cdot \frac{L^{d+1} \Gamma(\frac{1}{d}) \Gamma(r+1)}{d^2 \Gamma(r+\frac{1}{d}+2)} \\ &< \frac{\tau(d) \cdot \Gamma(\frac{1}{d}) L^{d+1} \cdot \Gamma(r+1)}{d^2 \Gamma(r+2)} \\ & (\because \Gamma(r+2) < \Gamma(r+\frac{1}{d}+2)) \\ &= \frac{\Gamma(\frac{1}{d})}{d^2 \sqrt[d]{\tau(d)}} \cdot \frac{N^{1+\frac{1}{d}}}{r+1} \quad (\because (4)) \end{aligned} \quad (20)$$

As a next step, let us consider the flooding strategy which can be considered as the one step ERS with SSV $\{L\}$. The expected cost of this strategy is given by,

$$\begin{aligned} C_{s,flt}^{(d)} &= C_f^{(d)}(L) = 1 + \tau(n)(L-1)^d \\ &\leq \tau(n)L^d = N \end{aligned} \quad (21)$$

If we apply Lemma 2 using Equation (20) and (21) we can conclude that the optimal expected search cost $C_{s,un}^{(d)}$ is $O(N/r)$. With the result of the lower bound of the optimal cost in Equation (19) we reasonably approximate that the optimal

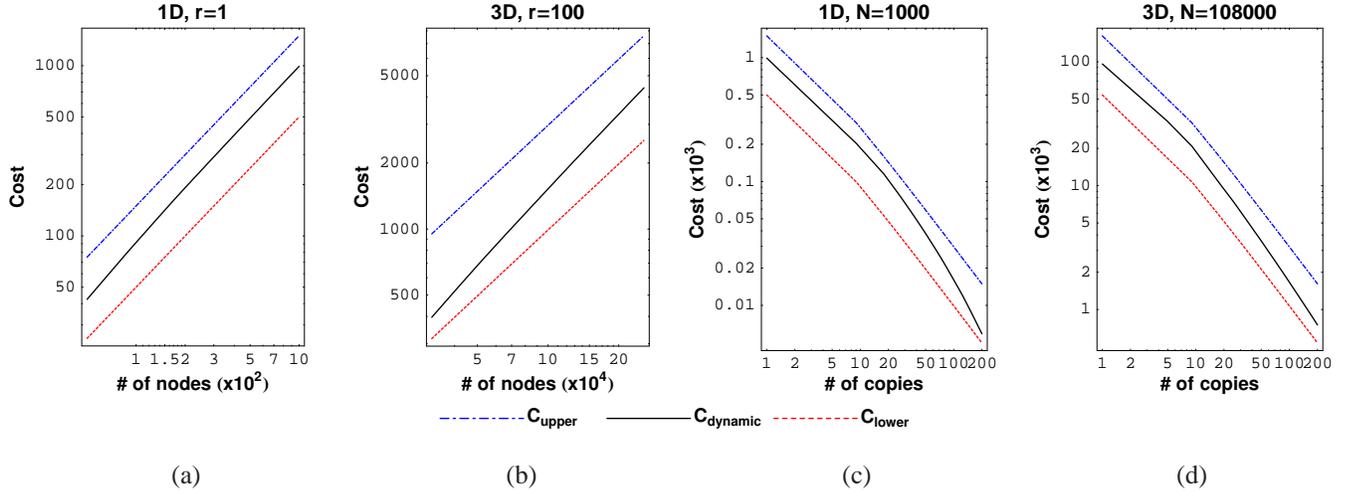


Fig. 2. Unstructured Average Search Cost

search cost is proportional to its corresponding lower bound. Hence, we have

$$C_{s,un}^{(d)} = \alpha_2 \cdot \frac{N}{r+1} \quad (22)$$

where α_2 is constant w.r.t r and N , but a function of d .

Figure 2 has similar plots as in Figure 1. The lower bound is using $C_{s,lower}^{(d)}$ in Equation (19), the upper bound is using $3C_{s,lower}^{(d)}$, and the search cost $C_{dynamic}$ is evaluated numerically using the dynamic programming algorithm proposed by Chang and Liu [6]. We can see that our approximation of the proportionality is quite close as for the structured search cost.

Lemma 2: Suppose $h_1(x) > 0$, $h_2(x) > 0$ for every $x > 0$, $g_1(y) > 0$, $g_2(y) > 0$ for every $y > 0$, and $f(x, y) > 0$ for every $x > 0$, $y > 0$. And suppose

- 1) $f(x, y) \leq \frac{h_1(x)}{g_1(y)}$ for every $x > 0$, $y > 0$
- 2) $f(x, y) \leq \frac{h_2(x)}{g_2(y)}$ for every $x > 0$, $y > 0$
- 3) $h_1(x) = O(x^n)$ and $h_2(x) = O(x^m)$, where $n > m > 0$
- 4) $g_1(y) = \Omega(y^a)$ and $g_2(y) = \Omega(y^b)$, where $a > b > 0$

Then,

$$f(x, y) = O\left(\frac{x^m}{y^a}\right) \quad (23)$$

Proof: From 3) and 4), there exist $c_1 > 0$ and $c_2 > 0$ such that

$$h_2(x) \leq c_1 \cdot x^m, \quad \text{for } \forall x \quad (24)$$

$$g_1(y) \geq c_2 \cdot y^a, \quad \text{for } \forall y \quad (25)$$

Suppose $f(x, y) = \Theta(x^{m+p} \cdot y^{-a+q})$ where $p, q \in \mathbb{R}$. Then, there exists $c > 0$ such that

$$f(x, y) \geq c \cdot x^{m+p} \cdot y^{-a+q}, \quad \text{for } \forall x, \forall y \quad (26)$$

If $p > 0$, then for every $y > 0$,

$$x^p > \frac{c_1}{cy^{-a+q}g_2(y)}$$

for sufficiently large x due to the archimedean property.

$$\Rightarrow f(x, y) \geq cx^{m+p}y^{-a+q} > \frac{c_1x^m}{g_2(y)} \geq \frac{h_2(x)}{g_2(y)} \quad (\because (24))$$

which is contradiction to 2). Hence, $p \leq 0$.

If $q > 0$, then for every $x > 0$,

$$y^q > \frac{h_1(x)}{cc_2x^{m+p}}$$

for sufficiently large y due to the archimedean property.

$$\Rightarrow f(x, y) \geq cx^{m+p}y^{-a+q} > \frac{h_1(x)}{c_2y^a} \geq \frac{h_1(x)}{g_1(y)} \quad (\because (25))$$

which is contradiction to 1). Hence, $q \leq 0$.

From $p \leq 0$ and $q \leq 0$, we conclude that $f(x, y) = O(x^m \cdot y^{-a})$. ■

V. COSTS UNDER UNIFORM RANDOM DEPLOYMENT (GRG MODEL)

We now consider the uniform random deployment for both structured and unstructured networks, i.e. the GRG model. As noted before, in order to maintain the connectivity under this deployment with high probability, the neighbor density τ should be logarithmically increasing with N . Hence, letting $\tau = \rho f(d)R^d = \theta_1 \log N$ with some constant θ_1 , the radio range can be expressed as,

$$R = \sqrt[d]{\frac{\theta_1}{\rho f(d)}} \sqrt[d]{\log N} \quad (27)$$

Note that our analysis in section III and IV is still valid except that the search costs derived therein are in terms of number of transmissions and the network radius L is now also a function of N . It should be noted that the number of transmissions is no longer proportional to the energy cost in the uniform random deployment because R is no longer constant with respect to N . Hence, we shall refer the derived search costs in the previous sections as H_{st} and H_{un} for structured and unstructured networks, respectively, in this section.

For structured networks, by substituting Equation (27) into (8) and (9), the bounds of H_{st} can be expressed as,

$$l'(d) \cdot \frac{\sqrt[d]{N}}{\sqrt[d]{r \log N}} < H_{st}^{(d)} < u'(d) \cdot \frac{\sqrt[d]{N}}{\sqrt[d]{r \log N}} \quad (28)$$

where

$$l'(d) = 2 \frac{\Gamma(\frac{1}{d}) \exp(\frac{1}{d})}{d \sqrt[d]{\theta_1}} \left(\frac{d}{d+1} \right)^{\frac{3d+2}{2d}}$$

$$u'(d) = 2 \frac{\Gamma(\frac{1}{d}) \exp(\frac{1}{d} + \frac{12+d}{12(12+13d)})}{d \sqrt[d]{e \theta_1}}$$

Now we introduce the transmission energy cost over a link of distance R given by,

$$E_t(R) = \beta R^\eta \quad (29)$$

where β is the transmit amplifier constant and η is the path-loss exponent. Generally, there should be a term for the distance-independent energy cost of transmitter and receiver electronics, but we assume it is negligible because we focus more on the search cost behavior of large networks¹.

Because the search cost is given by multiplying H_{st} and $E_t(R)$, we can approximate it using its double inequalities as follows:

$$C_{s,st}^{(d)} = \alpha_3 \cdot \frac{\sqrt[d]{N(\log N)^{\eta-1}}}{\sqrt[d]{r}} \quad (30)$$

where $\beta \left(\frac{\theta_1}{\rho f(d)} \right)^{\frac{\eta}{d}} \cdot l'(d) < \alpha_3 < \beta \left(\frac{\theta_1}{\rho f(d)} \right)^{\frac{\eta}{d}} \cdot u'(d)$.

With the analogous reasoning, we can obtain the search cost of unstructured networks under the GRG deployment as follows:

$$C_{s,un}^{(d)} = \alpha_2 \cdot \beta \left(\frac{\theta_1}{\rho f(d)} \right)^{\frac{\eta}{d}} \cdot \frac{N(\log N)^{\frac{\eta}{d}}}{r+1} \quad (31)$$

where α_2 is as same as in section IV.

We note that the search cost of unstructured networks under the GRG deployment turns out to be dependent on d , while that of the FTP deployment is independent of d .

VI. DISCUSSIONS AND CONCLUSIONS

We have derived minimum expected search energy costs for several kinds of data-centric wireless sensor networks. In particular, we have considered structured and unstructured networks deployed in d -dimensional area with a constant node density. In structured networks, the search cost of the FTP deployment of nodes is proportional to $\frac{\sqrt[d]{N}}{\sqrt[d]{r}}$, while that of the GRG deployment is proportional to $\frac{\sqrt[d]{N(\log N)^{\eta-1}}}{\sqrt[d]{r}}$. In unstructured networks, the search cost of the FTP deployment is proportional to $\frac{N}{r+1}$ regardless of the spatial dimension d of the network, while that of the GRG deployment is proportional to $\frac{N(\log N)^{\frac{\eta}{d}}}{r+1}$.

¹For more accurate result for small size networks, we can add a constant for the electronics energy cost to the transmission energy cost model, which would lead to the desired result without difficulty.

Although we have shown results in this paper assuming constant spatial node density and two specific cases of neighbor density scaling: constant and logarithmic, the approach presented here can be easily extended to consider other kinds of spatial and neighbor density scaling. For example, if we fix the network radius L and the radio range R and let the node density increase as N grows, it is easy to obtain that the search cost of structured networks under the FTP deployment is proportional to $\frac{L}{\sqrt[d]{r}}$ regardless of N , while that of unstructured networks is kept proportional to $\frac{N}{r+1}$.

One caveat to our models originates from the fact that we have resolved to approximate summations in both structured and unstructured cases using integrations since the summations fail to produce tractable closed-form expressions. When the radius of the network L is small or the number of copies r is very large compared to N , the approximations exhibit poor performance. The reason for the latter case is that larger r makes the curves of the integrands sharper.

We believe that the results presented here will be useful analytical tools for exploring the general performance of data-centric wireless sensor networks.

REFERENCES

- [1] S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, R. Govindan, L. Yin, and F. Yu, "Data-Centric Storage in Sensornets with GHT, A Geographic Hash Table", In *Mobile Networks and Applications (MONET), Special Issue on Wireless Sensor Networks*, 8:4, Kluwer, August 2003.
- [2] X. Li, Y.J. Kim, R. Govindan, and W. Hong, "Multi-dimensional Range Queries in Sensor Networks", *The First ACM Conference on Embedded Networked Sensor Systems (Sensys03)*, November 2003.
- [3] M.D. Penrose, "The longest edge or the random minimal spanning tree", *Annals of Applied Probability* 7, 340-361, 1997
- [4] M.D. Penrose, "A strong law for the longest edge of the minimal spanning tree", *Annals of Probability* 27, 246-260, 1999
- [5] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming*, W.M. McEneaney, G. Yin, and Q. Zhang, Eds. Boston, MA: Birkhauser, 1998, pp. 547-566.
- [6] N. Chang and M. Liu, "Revisiting the TTL-based Controlled Flooding Search: Optimality and Randomization," *Proceedings of the Tenth Annual International Conference on Mobile Computing and Networks (ACM MobiCom)*, September, 2004.
- [7] B. Krishnamachari, J. Ahn, "Optimizing Data Replication for Expanding Ring-based Queries in Wireless Sensor Networks", *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WIOPT '06)*, April 2006, Boston, USA.
- [8] J. Ahn and B. Krishnamachari, "Derivations of the Expected Energy Costs of Search and Replication in Wireless Sensor Networks", *USC Computer Engineering Technical Report CENG-2006-3*, April 2006.
- [9] J. Ahn and B. Krishnamachari, "Fundamental scaling laws for energy-efficient storage and querying in wireless sensor networks", *Proceedings of the Seventh ACM international Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '06)*, May 2006.
- [10] H. Robbins, "Remark of Stirling's Formula." *Amer. Math. Monthly* 62, 26-29, 1955.
- [11] J. Ahn and B. Krishnamachari, "Fundamental scaling laws for data-centric storage and querying in wireless sensor networks", *IEEE/ACM Transactions on Networking* in submission.