# Bounds on the Error in Estimating Redundancy in Randomly Deployed Wireless Sensor Networks

Renita Machado, Sirin Tekinay
ECE Dept., New Jersey Institute of Technology
Newark, NJ 07102
rmm23@njit.edu, tekinay@adm.njit.edu

## Abstract

**A challenging requirement in wireless sensor networks is the deployment of nodes in a wireless sensor network to satisfy continuous sensing with extended network lifetime while maintaining uniform coverage in the deployment region. While dense random deployments satisfy coverage and sensing requirements, constructing dense networks of sensor nodes poses economical constraints as well as the problem of redundancy. We provide an analytical framework for estimating the redundancy in a random deployment of nodes without the need of location information of nodes. We use an information theoretic approach to estimate the redundancy in a randomly deployed wireless sensor network and provide the Cramer-Rao bound on the error in estimating the redundancy in a wireless sensor network. We illustrate this redundancy estimation approach and calculate the bounds on the error in estimating the redundancy for a wireless sensor network with 1-redundancy. We also analytically show the inter-dependence between redundancy and network lifetime for random deployment.**

## 1. Introduction

Advances in miniaturization, embedded systems and wireless communications have significantly contributed to the success of WSNs for large-scale, low power networked sensing and data processing. A key challenge in WSN is the deployment of nodes to satisfy the requirements of higher network lifetime with continuous coverage of the deployment region for reliable sensing. While continuous coverage and connectivity lead to higher reliability of sensing operation, it takes a toll on the battery life of individual nodes and consequently reducing network lifetime. One way of balancing this tradeoff is deploying more sensors than are required to cover the deployment region. While this redundancy approach can be used to increase network lifetime by the use of sleep-scheduling and power-aware routing protocols, the economic constraints of deploying large number of nodes poses a limitation.

The problem of deployment has been widely studied in the context of providing uniform coverage, connectivity and redundancy while optimizing the number of sensor nodes required for the sensing operation [1, 2]. The random placement of nodes is preferred over the deterministic placement approach for remote and hostile environments, where it is not possible to place sensors in a particular pattern to cover the entire deployment region, for example, forest ecology sensing environments. Deployment for such applications is carried out by randomly scattering the nodes over the deployment region. While this approach has the advantage of eliminating the overhead of planning and deterministic placement, it also gives rise to the problem of not knowing the density and location of sensors in the deployment region. In addition, some network operations require higher density of sensors in regions of high-interest phenomenon. Since equipping nodes with location detecting GPS receivers is expensive and we do not know the precise location of sensors due to the random deployment, we need to be able to estimate the probability density function (PDF) of node distribution over the placement region and calculate the redundancy of nodes. The nodes can then be instructed to follow some degree of sleep-awake duty cycle to satisfy the requirements of sensing for that area of the deployment region while also contributing to higher network lifetime.

The problem we study can hence be modeled as follows: In a wireless sensor network of randomly deployed, stationary, power-limited, homogenous sensor nodes designed to be operational for at least time $T$, where we have no location information, how do we estimate the redundancy of nodes in the network? The time constraint $T$ ensures that all nodes in the network are designed to perform sensing and data processing without loss of battery energy for at least entire duration $T$, thus increasing reliability of system operation. One way of doing this would be intelligent processing at the nodes to discover the number of neighbors in their one-hop sensing range and calculating the redundancy. However, this approach, calls for higher number of transmissions at every node to discover the number of neighbors, which leads to increase in the depletion of battery energy for the energy-constrained nodes. Another way of estimating the redundancy would be from an analysis of the node deployment strategy. While this approach eliminates the need for sensors to discover the level of redundancy in their neighborhood, it results in greater variance of the error in estimating the redundancy due to the uncertainty of the deployment strategy information. In this paper, we propose an information theoretic approach to estimate the redundancy in a randomly deployed network. Every node transmits its node ID to the base station in the initial phase of the network operation. From these transmissions, the base station gathers the node IDs and the signal strength information. The signal strength data is used to estimate the PDF $p_X(x)$ of the node distribution over the deployment region using the MinMax measure [3]. We then decompose this PDF $p_X(x)$ into a primary distribution and an unknown number $\theta$ of secondary distributions. This unknown parameter $\theta$ is the redundancy that we estimate. Further, we provide bounds on the support of the random variable $x$ that describes the PDF of the node distribution over the deployment region. Finally, we provide the Cramer-Rao lower bound and an upper bound using Barron's proof [4] on the error in estimating the redundancy in the deployment region.

In [5], the authors study redundancy in terms of redundant broadcasts as a consequence of broadcasting by flooding in a mobile ad hoc network. By simulation, they show that for $k$ greater than or equal to four neighbors, the expected additional coverage is below 0.05 %, i.e. benefit of rebroadcast is small. In [6], the authors analyze sensor redundancy by finding bounds on the neighbor set of a sensor node. They provide an analytic framework to determine the percentage of redundant area with $n$ number of neighbors and the probability that a node is completely redundant. Specifically, they show that if a sensor is completely redundant, at least three and at most five neighbors are needed to cover its sensing area. The analytical model developed in [6] shows that for a 90 % partial redundancy, i.e. requiring 90% of its sensing area to be covered, needs five neighbors, which is similar to the simulation results obtained in [5]. Our work differs in that we do not require the sensors to be aware of the number of neighbors in their sensing range and thereby eliminate the need for processing and storage of the information related to neighbor discovery.

The rest of the paper is organized as follows: Section 2 presents the preliminaries for the WSN model and the PDF estimation technique based on the MinMax measure [3]. In section 3, we develop the analytical model to obtain the bounds on the Fisher information and the Cramer-Rao bounds for the error in estimating the redundancy parameter. We illustrate our approach with a case of 1-redundancy in the deployment region. Section 4 discusses the variation of error bounds of redundancy estimation and validates the relationship between redundancy and network lifetime. Section 5 concludes the paper and presents directions for future research.

## 2. Preliminaries

We assume a dense wireless sensor network of homogenous, stationary, power limited sensor nodes densely and randomly deployed over the deployment region designed to operate for at least $T$ time units. This time constraint on operation ensures that the resulting deployment and operation of the network is reliable for the desired interval. The problem is to estimate the redundancy in such a deployment without the knowledge of location information of sensor nodes. We assume that the base station initiates a phase of node-discovery, where it broadcasts a query transmission asking every node to respond to the base station with its node ID. From the signal strength of the transmissions of the nodes transmitting their node IDs, the base station obtains the relative location of the nodes in the deployment region. This problem is known as the *direct problem* [3], which refers to the problem of finding the initial probability assignment consistent with available information about a probabilistic system. In our problem formulation, the signal strength of individual transmissions comprises the available information about the node locations in the deployment region. The signal strength information constitutes the sample data and is used to estimate the PDF of node distribution over the deployment region. In this paper, we use the approach developed in [3] to obtain the PDF. We refer the reader to [3] for a detailed explanation of the approach. We proceed to estimate the redundancy of nodes in the deployment region from the PDF of the node location. We introduce the following definitions to aid in the redundancy calculations:

*Primary distribution:* The primary distribution refers to the PDF of sensor nodes' distribution that is necessary and sufficient to provide coverage and connectivity in the deployment region. The primary distribution ensures continuous sensing throughout the deployment region.

*Secondary distributions*: In order to provide continuous sensing in spite of battery exhaustion or device failure in nodes belonging to the primary distribution, as well as to improve the reliability of the sensing operation, we deploy additional nodes over the deployment region. The distribution of these redundant nodes has a PDF called the secondary distribution. Depending on the sensing requirements, minimum network lifetime constraints and the economic constraints and resources available for the sensing operation, we can deploy *k* distributions of sensor nodes in addition to the primary distribution over the deployment region resulting in *k*-redundancy.

The final step in the problem of redundancy estimation would be to estimate the number of secondary distributions (redundancy) over the deployment region from the initial sample set of signal strengths that provides the relative location of nodes in the deployment region. We also provide Cramer-Rao bounds on the error in estimating the redundancy in the deployment. To do this, we determine the Fisher information of the redundancy parameter θ, which is obtained from the estimated PDF (solution to the *direct problem*). This is a more general version of the redundancy estimation problem [6] in the absence of location information. Our results bring insight into the general problem of redundancy in dense, large WSNs where the estimate of redundant distributions can be useful to selectively power down or sleep schedule nodes in certain distributions to satisfy lifetime constraints or increase the density of sensing operation in high interest areas of the deployment region. The advantage of this approach lies in the fact that nodes do not need to possess computational complexity to process information about their neighbors to calculate redundancy in their coverage areas. The base station performs the processing to determine the redundancy in the entire deployment region. This approach can also be used to control the density of 'awake' nodes in sections of the deployment region by applying the redundancy processing approach to specific areas that exhibit high interest phenomenon.

## 3. Cramer-Rao bounds in error in redundancy estimation using the Fisher information of redundancy

To illustrate this method of estimating redundancy, we assume a primary distribution and a single secondary distribution of nodes over the deployment region, thus creating a 1-redundant network of nodes.

### A. Bounds on the PDF of node distribution over the deployment region

Let the primary distribution have a PDF $p_S(x)$ and let the secondary distribution have a PDF $p_Z(x)$, where $S$ is a random variable describing the primary distribution, and $X = S + Z_s^{(\tau)}$ describes the resultant distribution of all nodes in the deployment region. The secondary distribution $Z_s^{(\tau)}$ is a normal distributed random variable $\sim N(0,\tau)$ independent of *S*. Thus, $X$ which denotes the PDF of all nodes in the entire deployment region is a perturbed random variable with continuously differentiable density $p_X^x(\tau)$. Defining the score function as

$$\rho_X(x) = p_X^{'}(x) / p_X(x) \tag{1}$$

and $p^{(2\tau)}$ for the density of $S + Z_S^{(2\tau)}$, there exists a constant [4]

$$c_{\tau,k} = \sqrt{2}\left(2k/\tau e\right)^{k/2} \tag{2}$$

such that for all $x$ [4],

$$p_X^{(\tau)}\left(x\right)\left|\rho_X(x)\right|^k \le c_{\tau,k}\,p^{(2\tau)}(x) \qquad (3)$$

## B. Illustration of the redundancy estimation approach for 1-redundancy

A widely used assumption to model random deployment of nodes over a deployment region [1, 2] is the Poisson point process. Let the primary distribution $p_S(x)$ denote a Poisson point process of intensity $\lambda$.

$$\therefore\ p_S(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (4)$$

Since $X = S + Z_S^{(\tau)}$, $p_X^{(\tau)}\left(x\right) = p_S(x)\otimes p_Z(x)$ (5) Using Fourier transforms to obtain the convolution, the PDF $p_X^{(\tau)}\left(x\right)$ is given by

$$p_X^{(\tau)}\left(x\right) = \frac{1}{x!\,j^{(2x+2)}\,t^{x+2}}\sqrt{\frac{2\pi}{\tau}} \qquad (5)$$

The score function $\rho_X(x)$ for (5) is given by

$$\rho_X(x) = -\left[\frac{x\ln\left(-t\right)+1}{x}\right] \qquad (6)$$

The density $p^{(2\tau)}$ is obtained as follows

$$p^{(2\tau)} = \sqrt{\frac{\pi}{\tau}}\,\frac{1}{x!\,j^{2x+2}\,t^{x+2}} \qquad (7)$$

Substituting (6) and (7) in (3) and simplifying we get the bounds on the density function of the nodes in the deployment region,

$$p_X(x) \le \frac{-\sqrt{\frac{2\pi}{\tau}}\left(\frac{2k}{\tau e}\right)^{k/2}\frac{1}{x!\left(-t\right)^x t^2}}{\left[\frac{x\ln\left(-t\right)+1}{x}\right]^k} \qquad (8)$$

Substituting (5) in (8) to obtain the support for *x*, and evaluating *t* in the finite integral limits from 0 to $T$, where *T* is defined as the minimum time for which the network is designed to be operational (i.e. no node has run out of battery energy), the support for *x* is given by,

$$x \ge \frac{2}{\sqrt{\tau e}} - \ln\left(-T\right) \qquad (9)$$

Next, we evaluate the Fisher information of the redundancy parameter $\theta$. The Fisher information $J(X)$ is given by the variance of the score function and satisfies the following bound for the random variables *S* and *X* [4],

$$J(X) = E\left|\rho_X(x)\right|^2 \le \frac{5.658}{\tau e} \qquad (10)$$

Substituting for $\tau e$ from (9) in (10),

$$J(X) = E\left|\rho_X(x)\right|^2 \le \sqrt{2}\left(x+\ln(-T)\right)^2 \qquad (11)$$

In our problem formulation, the unknown parameter, which we estimate, is the redundancy $\theta$, the number of secondary distributions. To obtain the Fisher information of the redundancy parameter $J(\theta)$, we note that the Fisher information $J(X)$ can also be written as

$$J(X) = \int f(x-\theta)\left[\frac{\partial}{\partial x}\ln f(x-\theta)\right]^2 dx \qquad (12)$$

The Fisher information for the redundancy parameter

$$J(\theta) = \int f(x-\theta)\left[\frac{\partial}{\partial\theta}\ln f(x-\theta)\right]^2 dx \qquad (13)$$

Further, since $\frac{\partial}{\partial x}\ln f(x-\theta) = -\frac{\partial}{\partial\theta}\ln f(x-\theta)$,

$$\int f(x-\theta)\left[\frac{\partial}{\partial x}\ln f(x-\theta)\right]^2 dx = \int f(x-\theta)\left[\frac{\partial}{\partial\theta}\ln f(x-\theta)\right]^2 dx \qquad (14)$$

which implies $J(\theta) = J(X)$ (15)

The Cramer-Rao gives a lower bound on the estimation of variance of an unknown parameter. For a random variable *X* with mean $\mu$ and variance $\sigma^2$,

$$\sigma^2 \ge \frac{1}{J\left(X\right)}, \qquad (16)$$

with equality if and only if *X* is $N(\mu,\sigma^2)$.

The inverse of the Fisher information given by $1/J(X)$ is the Cramer-Rao bound. The Cramer-Rao inequality places a lower bound on the mean square error of an unbiased estimate of $X$ given by $\sigma^2$. Since $J(\theta) = J(X)$,

$$\frac{1}{\sigma^2} \le J(\theta) \le \sqrt{2}\left[x + \ln(-T)\right]^2 \tag{17}$$

The Cramer Rao bound on the error in estimating the redundancy $\theta$ for the case of a primary Poisson distribution and a single secondary normal distribution is thus given by

$$\frac{1}{\sqrt{2}\left[x + \ln(-T)\right]^2} \le \frac{1}{J(\theta)} \le \sigma^2 \tag{18}$$

**4. Results**

We now show the significance of (18) in establishing the bounds of error in redundancy estimation as a function of the expected network lifetime. We calculate the variance of $p_X(x)$ from (11) as

$$\sigma^2 = (-T)^{-\left(\frac{2}{\sqrt{\tau e}} - \ln(-T)\right)} \left[ \frac{-2K_1}{\sqrt{\tau e}} - \frac{K_2 (-T)^{-\left(\frac{2}{\sqrt{\tau e}} - \ln(-T)\right)}}{\left(\frac{2}{\sqrt{\tau e}} - \ln(-T)\right)^2} \right] \tag{19}$$

where

$$K_1 = \frac{-1}{T^2}\sqrt{\frac{2\pi}{\tau}}\ln(-T) \quad \text{and} \quad K_2 = \left[\frac{K_1}{\ln^2(-T)}\right]^2 \tag{20}$$

Evaluating (19) at the support of $x$ from (9), the bounds on the error in estimating redundancy are as follows,

$$\frac{\tau e}{4\sqrt{2}} \le \frac{1}{J(\theta)} \le \sigma^2 \tag{21}$$

Let $\tau$ be a fraction $\alpha$ of the desired network lifetime T. We show that the error bounds exist value as long as $\alpha \le$ constant $c_1$. For $c_1 \le \alpha \le c_2$, bounds may hold true depending on designed network lifetime T. For $\alpha > c_2$, there is no upper bound on the error information, i.e. the error in estimating the redundancy is infinite. These bounds are illustrated as shown in Figure 1. The concentric circles represent the areas where the bounds exist for corresponding values of $T$ and $\alpha$. As seen from Fig. 1, the circles are plotted in increasing order of desired network lifetime. We see that with an increase in network lifetime, the upper bound of the error in estimating the redundancy keeps decreasing. Since the lower bound in (21) is a constant, this implies that the error bound gap decreases for higher desired network lifetime. This can be intuitively explained as follows: For a larger desired network lifetime with given number of nodes, we need to maintain a lower value of redundancy in the deployment region. A lower value of redundancy translates to lesser number of 'awake' nodes covering the deployment region at any given time instant, thus increasing the network lifetime. Conversely, for lower desired network lifetime, higher number of nodes can be made to cover the deployment region by staying 'awake' for longer time intervals leading to a higher redundancy. Thus, the error in estimating a lower value of redundancy (higher network lifetime) is smaller than that in estimating a higher value of redundancy (lower network lifetime). Numerical results from simulating (21) for different desired network lifetimes T, shows that for $\tau \le 0.7$, the error is bounded. The upper bound of error for $T$=1000 hours and 100 hours are close and shown as overlapping circles ($\tau \le 0.01$). The next larger circle shows the error bound for $T$=10 hours with $\tau \le 0.03$. The shaded region in Fig.1 represents the value of lower network lifetimes for which the error bounds may not be satisfied due to decrease in *T.* For extremely low value of desired network lifetime, the error bounds for redundancy estimation are not satisfied, implying that the error in estimating redundancy is infinite. Fig.2 shows the upper bound of error estimation as a function of support of the random variable representing the node distribution for 1-redundancy with T = 5 hours. We see that for increasing lower limit of the support, the error in estimating the redundancy increases due to the inverse dependence of support on *T*.

## 5. Conclusions

The paper presented an analytical model to estimate the redundancy in a randomly deployed WSN without the use of location information. We illustrated the redundancy calculation

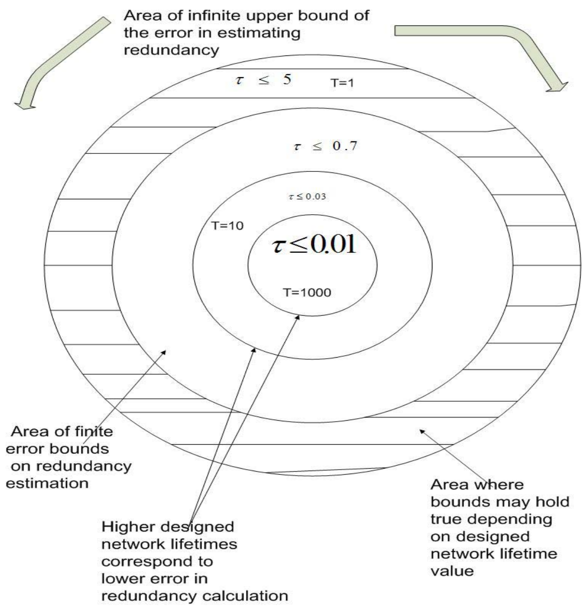


Fig.1. Interdependence of network lifetime and error in redundancy estimation for 1-redundancy. Higher network lifetimes correspond to lower redundancy and hence lower error in estimating redundancy. For the converse case of lower network lifetime, the error in redundancy estimation is higher.

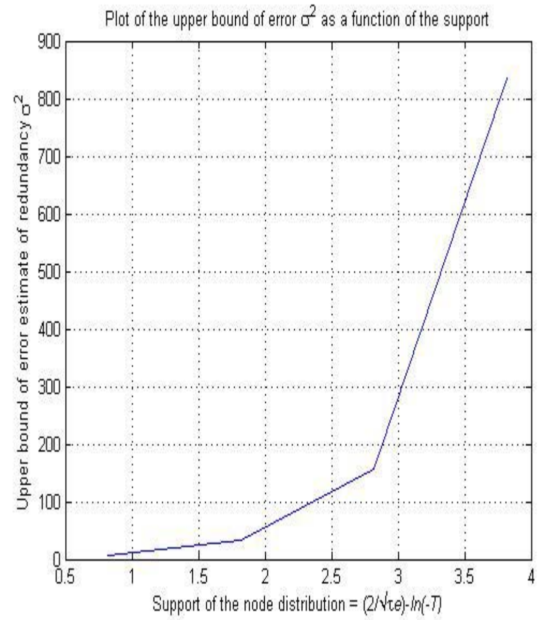


Fig.2. Variation of error bound gap as a function of support of the random variable describing node distribution. The support is inversely dependent on the network lifetime. A lower value of the support implies a higher network lifetime and hence lower error in redundancy and its estimation.

approach and obtained the bounds on the error in estimating the redundancy for 1-redundancy in a randomly deployed WSN. We also showed the mutual dependence of network lifetime and redundancy in node deployment. These results can be used to design efficient sleep scheduling mechanisms to improve network lifetime by selectively controlling the density of 'awake' nodes in the deployment region. Further, the use of redundancy information can be employed to increase the reliability of sensing operation for regions of high-interest phenomenon that require larger number of sensors to accurately sense the environment. Our future work would involve analyzing the redundancy problem with varying degrees of redundancy and different node distributions. A simulation model to obtain PDF of nodes from a randomly deployed WSN and then estimating the bounds on the error in redundancy would be the next step in this research.

## 6. References

[1] P. Wan and C. Yi, "Coverage by Randomly Deployed Wireless Sensor Networks", *Fourth IEEE International Symposium on Network Computing and Applications*, 2005, pp. 275-278.

[2] B.Wang, K.C. Chua, V.Srinivasan, W.Wang, "Sensor Density for Complete Information Coverage in Wireless Sensor Networks", in *EWSN 2006, pp.69-82*.

[3] S. Munirathnam, H.K. Kesavan, and P.H. Roe, "Probability density function Estimation using the MinMax Measure", in *IEEE Transactions on Systems, Man and Cybernetics, Part C: applications and Reviews*, Volume 30, Issue 1, Feb 2000, pp 77-83.

[4] Oliver Johnson, *Information Theory and the Central Limit Theorem*, Imperial College Press, London, 2004.

[5] S. Ni, Y. Tseng, Y. Chen and J. Sheu, "The broadcast storm problem in a mobile ad hoc network", in *Proceedings of the 5th Annual International Conference on Mobile Computing and Networking (MobiCom 99)*, Seattle, USA, August 1999, pp. 151-162.

[6] Y. Gao, K. Wu and F. Li, "Analysis of the redundancy of wireless sensor networks", in *Proceedings of the 2nd ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)*, San Diego, CA, September 2003.