

# SPATIAL CORRELATION BASED SENSOR SELECTION SCHEMES FOR PROBABILISTIC AREA COVERAGE

Ramesh Rajagopalan

School of Engineering, University of St. Thomas, MN, USA

ramesh@stthomas.edu

## **ABSTRACT**

*This paper develops an analytical model for probabilistic area coverage in terms of the target detection probability. A decision fusion framework is utilized to infer the presence or absence of the target. Analytical results are derived for the target detection and false alarm probabilities in the presence of correlated sensor noise. The spatially correlated sensor observations are utilized to select a subset of sensors to meet the specified area coverage. Two new sensor selection schemes are proposed for maximizing information theoretic measures such as joint entropy. The sensor selection schemes are analyzed extensively based on simulations. The results show that the proposed sensor selection scheme outperforms two state-of-the-art sensor selection schemes: constrained random sensor selection and disjoint random sensor selection.*

## **KEYWORDS**

*Probabilistic area coverage, spatial correlation, target detection, sensor selection, joint entropy*

## **1. INTRODUCTION**

Wireless sensor networks (WSNs) have been used for numerous applications such as military surveillance, environmental monitoring, intrusion detection and smart homes. In many of these applications, the sensors observe the physical space or phenomenon of interest, and report data to the fusion center. However, practical considerations require optimal use of limited resources such as bandwidth, energy, storage and computation/communication capabilities, while satisfactorily covering the region being monitored. Hence this paper addresses the probabilistic area coverage problem, in which it is necessary to determine a minimal subset of sensors sufficient to meet the prespecified area coverage criterion, using a probabilistic sensing model. In order to reduce (i) the amount of energy used for communication, and (ii) traffic flow to the fusion center (avoiding network congestion), an appropriate data delivery model must be chosen. Three important data delivery models are: continuous, event driven and query based [1]. In the continuous model, all sensors transmit their data periodically to the fusion center while in the event driven model, sensors report their sensed data only when a specific event occurs. In this paper, a query based data gathering model is considered where the fusion center periodically queries the sensors for data.

Selection of sensors to be queried should be based on the fact that the observations of sensors in close spatial proximity are highly correlated, especially in densely deployed sensor networks. The degree of correlation increases with decrease in inter-sensor separation. Such spatial correlations can be exploited: the fusion center should collect data from well separated sensors, reducing the redundancy between the data collected from different sensors. Correlation among sensor observations can be caused by colored observation noise which arises in densely

deployed sensor networks. In many applications, the phenomenon to be measured by the sensors is subject to correlated ambient noise.

This paper addresses two important questions for the area coverage problem:

- What is the relationship between the number of sensors ( $M$ ) and the probabilistic area coverage?
- To meet a desired coverage, which  $M$  sensors should the fusion center select in each communication round?

In order to answer these questions, the area coverage properties of sensor networks are investigated under the probabilistic sensing model. The main contributions of this paper are:

- Development of a new model to quantify the area coverage of the sensor network in terms of real world applications such as target detection.
- Development of new sensor selection schemes which select the most informative subset of sensors.

The rest of the paper is organized as follows. Section 2 presents recent work on area coverage and sensor selection problems. Section 3 describes the area coverage problem and characterizes the area coverage in terms of distributed target detection. Section 4 formulates the sensor selection problem and describes the spatial correlation model. Section 5 describes different sensor selection schemes.

Section 6 analyzes simulation results and Section 7 summarizes the contributions.

## **2. RELATEDWORK**

### **2.1. Area Coverage Problem**

Coverage problems in sensor networks have been an active research area in the recent past. In area coverage problems, the main goal of the sensor network is to collectively monitor or cover a region of interest. Several algorithms have been developed for achieving the desired area coverage. Tian and Georganas [2] have proposed a node scheduling scheme to meet the area coverage requirement and to extend the lifetime of the network. Their scheme conserves energy by turning off the redundant sensors in the areas fully covered by the other sensors while guaranteeing area coverage.

The coverage of a grid-based wireless sensor network has been studied by Shakkottai et al. [3]. They have considered the coverage and connectivity problem and derived necessary conditions on the sensing range and failure rate of the sensors to achieve the desired coverage. Choi et al. [4] have developed an energy conserving data gathering approach which characterizes the tradeoff between coverage and data reporting latency.

S. Slijepcevic, and M. Potkonjak [5], and Cardei et al. [6] aim to achieve energy efficient area coverage by dividing the nodes in the sensor network into disjoint sets such that every set can uniquely provide area coverage. The sets are activated successively for area coverage. When a particular set is active, the other sets are in a sleep mode.

In the point coverage problem, the objective is to cover a set of points in an area. Cardei and Du [7] address the point coverage problem in which a limited number of points with known locations need to be monitored. The proposed solution models disjoint sets as disjoint set covers such that every cover completely monitors all target points. The disjoint set cover

problem is modeled as a mixed integer programming problem and solved using a polynomial time approximation algorithm. Liu et al. [8] have studied the coverage properties of large scale sensor networks. They have studied three fundamental coverage measures: area coverage, node coverage and detectability under the Boolean and the probabilistic sensing models. A recent survey on energy efficient area coverage problems is presented in [9].

Most of the work in area coverage problems mentioned above assume a binary sensing model in which the sensor readings have no associated uncertainty. These models assume that a sensor detects any object within its sensing range with probability 1. However the detection probability of a sensor depends on the distance between the sensor and the object within its sensing range. In addition, the sensor readings are uncertain and the coverage needs to be expressed in probabilistic terms. This paper addresses these issues by developing a new area coverage metric that quantifies the target detection and false alarm probabilities. In addition, this work also integrates the spatial correlation of sensor observations in to the distributed detection problem. Analytical results are derived for the target detection and false alarm probability in the presence of correlated Gaussian noise.

## **2.2. Sensor Selection Problem**

In many sensor network applications, a selected set of sensors report data to the fusion center to conserve energy and bandwidth. Several algorithms have been proposed in the literature to solve the sensor selection problem. Wang et al. [10] have proposed an entropy based sensor selection heuristic for target localization problems in sensor networks. They have proposed a greedy sensor selection heuristic that selects the most informative sensor in each round that would yield the greatest reduction in the entropy of the target location distribution. Choi et al. [11] have proposed a constrained random sensor selection scheme (CROSS) which employs a Poisson sampling technique that maintains a certain minimum distance between the sampled sensors. The main goal of CROSS is to maximize the area coverage and reduce the overlap between sensing ranges of selected sensors. Their simulation results show that CROSS performs better than the disjoint random sensor selection scheme. Doherty et al. [12] have considered the scattered data selection for field reconstruction problem. They have proposed different sensor selection schemes such as hop based selection and extremum selection aimed at optimizing the energy efficiency of the network.

The main drawback of the above mentioned schemes is the lack of a spatial correlation model, essential to capture the redundancy of the data collected by different sensors. The main goal of sensor selection is to maximize the information extracted from the selected sensors. Hence, a good sensor selection scheme should optimize information theoretic measures such as joint entropy of the observations collected from multiple sensors. This paper attempts to address these issues by developing sensor selection schemes based on an underlying spatial correlation model. Unlike prior work in area coverage problems, our approach solves the joint problem of sensor selection and area coverage under a probabilistic sensing model.

## **3. PROBABILISTIC AREA COVERAGE**

This section describes the distributed target detection problem in the presence of correlated Gaussian noise. A decision fusion framework is utilized to solve the detection problem at the fusion center. Analytical expressions are derived for the target detection and false alarm probabilities and a new definition of area coverage is presented.

### 3.1. Location Model

Consider a sensor network in which a total of  $N$  sensors are randomly deployed in a two dimensional geographical region. In harsh and unfriendly environments such as a battlefield, the sensors may be air-dropped. The region of interest is modeled as a square with area  $a^2$ . The location of sensors are i.i.d and follow a uniform distribution given by

$$f(x_i, y_i) = \frac{1}{a^2} \quad (0 \leq x_i, y_i \leq a) \quad (1)$$

### 3.2. Sensor Measurement Model

In the binary sensor model, a sensor detects an object within its detection range with probability 1, i.e., the sensor readings have no associated uncertainty. In addition, the binary model does not capture the degradation of a sensor's sensing capability with the increase in distance between the sensor and the target. By contrast, this paper uses a realistic sensing model which explicitly characterizes the uncertainty in detection by any sensor. In this model, sensors perform detection by measuring the signal power emitted by the target. The energy of physical signals such as acoustic and electromagnetic signals attenuates with the distance from the signal source. Let  $d_i$  indicate the distance between sensor  $i$  and the target. The attenuated signal power  $s_i$  is given as

$$s_i = P_0 \cdot g(d_i) \quad (2)$$

where  $P_0$  is the power emitted by the target and  $g(\cdot)$  is a decreasing function satisfying  $g(0)=1$ ,  $g(\infty)=0$  and  $g(x)=\phi(x-k)$ . In this work, an isotropic signal attenuation model is adopted where the signal decay function is modeled as

$$g(d_i) = \frac{1}{1 + d_i^n} \quad (3)$$

where  $d_i$  is the distance between the target and local sensor  $i$ :

$$d_i = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \quad (4)$$

and  $(x_t, y_t)$  are the co-ordinates of the target. The signal decay exponent  $n$  can take values between 2 and 3. In this work,  $n$  is set to 2.

### 3.3. Distributed Target Detection

The distributed detection problem has been studied widely for the case where the observations are conditionally independent given either hypothesis (target present or absent). There has not been much work in the area of distributed detection in the presence of correlated sensor measurements. In [13], the authors have solved the problem of detecting a constant signal in the presence of correlated zero mean Gaussian noise. In this section, the results in [13] are used to obtain new analytical expressions for the target detection and false alarm probabilities.

Depending on the hypothesis that the target is absent ( $H_0$ ) or present ( $H_1$ ), the measurement at sensor  $i$  denoted as  $y_i$  is given as,

$$\begin{aligned} H_0 : y_i &= n_i \\ H_1 : y_i &= s_i + n_i \end{aligned} \quad (5)$$

where  $s_i$  is the signal power defined in equation (2) and  $\{n_i\}$  are dependent zero mean Gaussian noise with unit variance. In our model, the correlation coefficient between any two sensor observations decreases geometrically with increase in distance between the sensors. This model has (1) been widely adopted in the literature to capture the redundancy in the observations of spatially close sensors [13-15]. The observation noise at any two sensors  $i$  and  $j$  has a covariance function given by

$$\rho_{ij} = \sigma^2 \rho^{d_{ij}}; \quad 0 \leq \rho < 1 \quad (6)$$

where  $\rho$  is the correlation coefficient,  $d_{ij}$  is the distance between sensors  $i$  and  $j$ , and  $\sigma$  is the standard deviation of the noise which is assumed to be unity.

Each sensor compares the measurement  $y_i$  to a threshold  $\tau$  to make a decision about the presence or absence of the target. The false alarm probability at a local sensor is

$$\begin{aligned} p_{fa} = \text{prob}(y_i > \tau | H_0) &= \int_{\tau}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= Q(\tau) \end{aligned} \quad (7)$$

or

$$\tau = Q^{-1}(p_{fa}) \quad (8)$$

where  $Q(\cdot)$  is the complementary distribution function of the standard Gaussian:

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (9)$$

The probability of detection at sensor  $i$  is given as

$$P_{di}(x_i, y_i, x_t, y_t) = \text{prob}(y_i > \tau | H_1) = \int_{\tau}^{\infty} \left( \frac{1}{\sqrt{2\pi}} \right) e^{-\frac{(t-s)^2}{2}} dt = Q\left( \tau - \frac{P_0}{1+d_i^n} \right) \quad (10)$$

### 3.4. Decision Fusion

In decision fusion, each sensor compares its measurement to a threshold to arrive at a local decision about the presence of a target. The sensors transmit the local decisions to the fusion center. The fusion center compares the sum of local decisions to another threshold to arrive at a global decision.

This model is widely used in the literature to compute the global target detection probability [16, 17]. We denote the binary data sent from local sensor  $i$  as  $I_i, i=1,2,\dots,N$ . Since the local sensors transmit only binary data, the fusion center does not have the knowledge of detection probabilities at local sensors. In the absence of this information, the fusion center treats the decisions from the sensors equally and compares  $\Lambda = \sum_{i=1}^N I_i$  to a threshold  $\gamma$  to arrive at a global decision.

### 3.4.1. Calculation of $P_f$

The probability of false alarm at the fusion center is given by  $P_f = \text{Prob}[\Lambda \geq \gamma | H_0]$ . Since the local decisions at the sensors are correlated, analytical expression for the distribution of  $\Lambda$  is intractable. The results derived in [13] can be used to approximate the distribution of  $\Lambda$ . A central limit theorem for a sequence of correlated random variables in which the past and distance future (10) are asymptotically independent is described in [13]. According to our correlation model in equation (6),  $\rho_{ij} \rightarrow 0$  as  $d_{ij} \rightarrow \infty$ . Hence, the central limit theorem described in [13] can be used to approximate  $\Lambda$  using a Gaussian distribution. Assuming that the number of sensors is large,  $\Lambda$  follows a Gaussian distribution with mean  $Np_{fa}$  and variance  $\sigma_{N0}^2$ . Utilizing the results in [13] for  $\sigma_{N0}^2$  we have,

$$\sigma_{N0}^2 \cong N \left( \frac{1+\rho}{1-\rho} \right) [p_{fa} - p_{fa}^2] \tag{11}$$

Hence, the global probability of false alarm is given as

$$P_f = \text{Prob}[\Lambda \geq \gamma | H_0] = Q \left( \frac{\gamma - Np_{fa}}{\sigma_{N0}} \right) \tag{12}$$

### 3.4.2. Calculation of $P_d$

The detection probabilities at the local sensors are correlated and not identical since  $p_{di}$  at sensor  $i$  is a function of the distance  $d_i$  as shown in equation (10). If the number of sensors are sufficiently large, the results derived in [13] show that  $\Lambda$  follows a Gaussian distribution with mean  $\mu$  and variance  $\sigma_{N1}^2$  where

$$\mu = E[\Lambda] = \sum_{i=1}^N E[I_i] = \sum_{i=1}^N p_{di} \tag{13}$$

$$\sigma_{N1}^2 \cong \left( \frac{1+\rho}{1-\rho} \right) \sum_{i=1}^N p_{di} (1 - p_{di}) \tag{14}$$

Since,  $p_{di}$  is a function of the sensor and target locations, when  $N$  is large, the summation in equation (13) can be approximated as

$$\begin{aligned} \mu(x_t, y_t) &= \sum_{i=1}^N p_{di}(x_i, y_i, x_t, y_t) \\ &\approx \sum_{i=1}^N E\{p_{di}(x_i, y_i, x_t, y_t)\} \\ &= Np_{dav}(x_t, y_t) \end{aligned} \tag{15}$$

The sensor locations follow a uniform distribution as described in equation (1). Hence, we have

$$P_{dav} = \frac{1}{a^2} \int_0^a \int_0^a P_d(x, y, x_t, y_t) dx dy \quad (16)$$

where

$$P_d(x, y, x_t, y_t) = Q\left(\tau - \frac{P_0}{1 + ((x-x_t)^2 + (y-y_t)^2)^{n/2}}\right)$$

Similarly, the summation in equation (14) can be approximated as

$$\begin{aligned} \sigma_{N1}^2(x_t, y_t) &\approx N\sigma_{av}^2(x_t, y_t) \\ &= \frac{N(1+\rho)}{a^2(1-\rho)} \int_0^a \int_0^a (1 - P_d(x, y, x_t, y_t) \times P_d(x, y, x_t, y_t)) dx dy \end{aligned} \quad (17)$$

The global detection probability at the fusion center is given by

$$P_d(x_t, y_t) = \text{prob}(\Lambda > \gamma) \approx Q\left(\frac{\gamma - \mu(x_t, y_t)}{\sigma_{N1}(x_t, y_t)}\right) \quad (18)$$

The location of the target  $(x_t, y_t)$  is a random variable. Under the assumption that there is no prior information about the likelihood of the target location, a uniform distribution captures the uncertainty in the target location. Hence the average global detection probability at the fusion center is given by

$$P_{dav} = \frac{1}{a^2} \int_0^a \int_0^a P_d(x_t, y_t) dx_t dy_t \quad (19)$$

### 3.5. Area Coverage

In surveillance applications, the probability of false alarm together with the target detection probability characterizes the sensing quality provided by the network. For large scale networks, the area coverage can be characterized in terms of the detection probability. This paper presents a new concept of coverage called the  $(\epsilon, \lambda)$  coverage.

**Definition:**  $(\epsilon, \lambda)$  coverage: Given two constants  $\epsilon$  and  $\lambda$ , a target appearing at a point  $p$  in the surveillance region is covered if the global detection probability and false alarm probability at point  $p$  satisfy,

$$P_{dav} \geq \lambda \text{ and } P_f = \epsilon \quad (20)$$

Mission critical applications require a very low false alarm rate and a high detection probability. For instance, the coverage requirements can be specified as  $\epsilon=5\%$  and  $\lambda=80\%$ .

The coverage is closely related to the number of sensors in the network. For a given  $\epsilon$  and local sensor false alarm probability  $p_{fa}$ , the threshold  $\gamma$  can be computed using equation (12). We can then compute the global detection probability using equations (18) and (19).

#### 4. THE SENSOR SELECTION PROBLEM

If the desired  $(\epsilon, \lambda)$  coverage is specified by a design expert, we can compute the required number of sensors ( $M$ ) which should report data employing the analytical relation derived in the previous section. For the sensor selection problem, we consider large-scale sensor networks in which the number of sensors deployed  $N \gg M$ . However,  $M$  sensors are sufficient to achieve the required coverage. Now, we are ready to answer another important question: which  $M$  sensors should report data in each data-gathering round? This section addresses this issue by formulating and solving the sensor selection problem.

Given a set of sensors  $A = \{S_1, S_2, \dots, S_N\}$ , the goal of sensor selection is to select the most informative subset of sensors of cardinality  $M$  in each communication round. Hence the objective is to obtain a set of sensors  $B^*$  of cardinality  $M$  such that

$$B^* = \underset{B \subset A}{\operatorname{argmax}} H(B) \quad (21)$$

where  $H(B)$  denotes the joint entropy of the variables in  $B$ . In dense sensor networks, the data collected by sensors in close spatial proximity are strongly correlated with each other. The spatial correlation model described in equation (6) can be used to design a strategy for selecting sensors which are far apart from each other. The next section describes the information theoretic measure used to characterize the quality of data collected by the sensors.

##### 4.1. Differential Entropy

Differential entropy has been widely used to characterize the total amount of uncorrelated data generated by a set of sensors [14], [17-21]. A measure of the total amount of information gathered by the sensors is given by the differential entropy of the multivariate Gaussian distribution. According to the target detection model discussed in section 3.3, the measurements at different sensors  $Y = [y_1, y_2, \dots, y_N]$  form a multivariate Gaussian distribution. For an  $M$ -dimensional multivariate Gaussian random variable  $Y$ , the differential entropy is

$$h(Y) = \frac{1}{2} \times \log(2\pi e)^M \times \det(K) \quad (22)$$

where  $\det(K)$  indicates the determinant of the covariance matrix. According to the correlation model defined in equation (6), the noise covariance matrix is given as

$$K = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1N} \\ \rho_{21} & 1 & \dots & \rho_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{N1} & \rho_{N2} & \dots & 1 \end{bmatrix} \quad (23)$$

We assume that the noise variance is unity,  $N$  is the total number of sensors and  $\rho_{ij}$  is the covariance of measurement noise at sensors  $i$  and  $j$  defined in equation (6).



The information transmitted by the sensors to the fusion center is dependent on the specific application in hand. For instance, in acoustic target detection and localization, sensors transmit acoustic samples and the extracted target signature to the fusion center. We assume each sensor quantizes its samples independently into a  $g$  bit data packet. The discrete joint entropy associated with the quantized samples of  $M$  sensors is given by

$$H(Y)=h(Y)+gM \quad (24)$$

Since transmission of raw data consumes bandwidth and energy, it is imperative to select the most informative sensors for data transmission. The next section describes two new sensor selection schemes for optimizing the information theoretic metric – joint entropy defined in equation (24).

## 5. SENSOR SELECTION SCHEMES

In this section, two new sensor selection schemes are described: distance based sensor selection and entropy based sensor selection. The extent of correlation in the data observed by different sensors is a function of the distance between them. Intuitively, sensors that are far apart will have maximum uncorrelated information about the target. Hence, the distance based selection scheme selects sensors which are as far apart from each other as possible. The entropy based sensor selection scheme instead iteratively selects a sensor that can add the maximum information, given the sensors selected so far.

In both the sensor selection schemes, the fusion center selects the first sensor as the one with highest detected signal power, which is in close proximity to the target. This approach is also very useful in dynamic problems such as target tracking where the sensor measurement information such as location and velocity of the target changes with time. Hence, at each sensor, the amount of detected signal power from the target varies with time. The sensors transmit the detected signal power (not the raw measurement data) to the fusion center as a pre selection step. This avoids transmission of raw data such as target signatures to the fusion center and therefore saves limited bandwidth and prolongs battery lifetime for the sensors. After selecting the first sensor, the two selection schemes differ in the heuristic employed for selecting the remaining  $M-1$  sensors. It is interesting to note that two sensors in close spatial proximity could have high detected signal power if both sensors are close to the target. Hence the detected signal power alone cannot characterize the amount of uncorrelated data in the sensors. The proposed schemes are centralized where the fusion center performs the necessary computations to select the sensors. The sensor selection schemes are based on the following assumptions:

- $N$  sensors are deployed randomly (uniform distribution) in the sensor field.
- The fusion center has no energy constraints and has the pre-determined knowledge of the geographical locations (eg: through GPS interfaces) of all the sensors.

Although the proposed sensor selection schemes are designed for area coverage, they can be readily extended to dynamic problems such as target localization and tracking. In such cases, the entropy based information measure is calculated using the expected posterior distribution of the state to be estimated. The proposed sensor selection schemes can be used to query sensors for detailed information about the target such as target signatures and direction of arrival information. The sensor selection schemes select  $M$  sensors in each communication round.

Depending on the application at hand, the number of communication rounds can be determined by a higher level quality of service parameter such as the estimation error.

### 5.1. Distance Based Sensor Selection Scheme

The *Max-Min* distance based sensor selection scheme is described below in Figure 1; where  $F_c$  denotes the fusion center, and  $M$  denotes the number of sensors to be chosen.

1. **Transmission:** Every sensor node transmits its detected signal power to  $F_c$
2. **Selection of first sensor:**  $F_c$  selects the sensor  $S^*$  with the highest detected signal power.
3. **Initialization:** Let  $A$  = set of all sensors except  $S^*$ ,  $B = \{S^*\}$ .
4. **Iteratively choose other sensors:**
  - While*  $|B| < M-1$ , *do*:
  - a.  $D=0$ ;
  - b. *For each sensor*  $S_j \in A$ , *do*:
    - $d_j = \min$  [Euclidean distance from  $S_j$  to any element in  $B$ ];
    - if  $d_j > D$ , then  $S_{\#} = S_j$ ;
    - $D = d_j$ ;
  - c. Move  $S_{\#}$  from  $A$  into  $B$ .

Figure 1. Algorithm describing the distance based sensor selection process in a communication round

### 5.2. Entropy Based Sensor Selection Scheme

The goal of the entropy based sensor selection algorithm is to select  $M$  sensors with maximum discrete joint entropy in each communication round. In each round, a greedy approach is used to select  $M$  sensors. Figure 2 describes the algorithm, differing from the previous algorithm in Step 4. The discrete joint entropy is computed using equation (24).

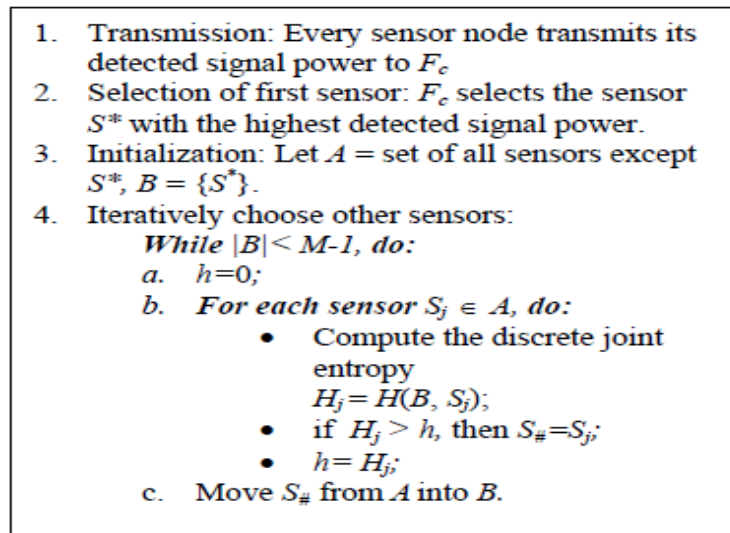


Figure 2. Algorithm describing the entropy based sensor selection process in a communication round

The greedy heuristic selects the next sensor as the one which has maximum discrete joint entropy with the sensors selected so far. This approach can be shown to have an approximation property, as described below. Functions which satisfy the diminishing returns property, defined as  $F(B \cup X) - F(B) \geq F(A \cup X) - F(A)$  ;  $B \subset A$ , are called submodular functions as described in [21]. The discrete joint entropy is a strictly non-decreasing submodular function which can effectively measure the total information contained in a set of random variables. We now summarize the theoretical results in [22] about a greedy algorithm for optimization of non-decreasing submodular functions.

1. The greedy algorithm guarantees an  $(1-1/e)$  approximation of the optimal solution with high confidence. In other words, the greedy algorithm obtains a set that is at most  $(1-1/e)$  worse than the optimal set.
2. There exists no polynomial time algorithm that can provide an approximation better than  $(1-1/e)$  unless  $\mathbf{P}=\mathbf{NP}$ .
3. The greedy algorithm obtains a set  $B \subseteq A$  of size  $M$  with computational complexity of  $O(M|A|)$  function evaluations.

Since discrete joint entropy is a non-decreasing submodular function, the entropy based sensor selection scheme guarantees a  $(1-1/e)$  approximation of the optimal solution with a computational complexity of  $O(M|A|)$ .

### 5.3. CROSS and DRS

We compare our new sensor selection schemes with two state of the art sensor selection schemes described in [11]: constrained random sensor selection (CROSS) and disjoint random sensor selection (DRS). The CROSS scheme randomly selects  $M$  sensors in each data gathering round while maintaining a minimum distance between any pair of selected  $M$  sensors. The desired minimum distance is computed based on the assumption that sensors detect the target within their sensing range with probability 1. In our work, the sensing model is probabilistic.

Hence, to adapt CROSS for our model, we approximate the sensing range  $d_{min}$  as the distance where the detection probability at a sensor is 0.9. The desired minimum distance  $d_{min}$  is then computed using equation (10). Hence, the selected  $M$  sensors cover the sensing area filled with non-overlapping circles with radius  $d_{min}$ .

In DRS scheme, the fusion center randomly selects  $M$  sensors in each data reporting round. DRS maintains a reporting cycle  $C$  which indicates the periodicity of a sensor to report data to the fusion center. The reporting cycle  $C$  contain  $\delta = \lfloor \frac{N-1}{M} \rfloor$  reporting rounds. The fusion center randomly selects a reporting round for each sensor from  $\delta$  rounds within  $C$  and keeps the selected round as the reporting round for that sensor.

## 6. SIMULATION RESULTS

We used Matlab for our simulations. Our sensor selection algorithms involve computation of joint entropy which requires evaluation of the determinant of the covariance matrix. These computations are well supported in Matlab environment. We also implemented CROSS and DRS schemes in Matlab. Extensive Matlab simulations were performed on a sensor field of area  $25 \times 25$  m<sup>2</sup>. The sensors were randomly deployed with uniform distribution as described in equation (1). The rest of the section is organized as follows. Section 6.1 analyzes the variation of area coverage with the number of sensors. Section 6.2 presents the performance comparison of different sensor selection schemes for different values of the correlation coefficient.

### 6.1. Analysis of Area Coverage

This section analyzes the impact of the number of sensors on the  $(\epsilon, \lambda)$  coverage. For the target detection model, we set  $n=2$  and  $P_0=100$ . We assume that the noise variance is unity and the correlation coefficient  $\rho=0.5$ . Figure 3 shows the variation of area coverage (detection probability  $\lambda$ ) with the number of sensors for  $\epsilon=5\%$ . The figure shows that the detection probability increases with the number of sensors. When the number of sensors is large, the detection probability converges to 1. If the desired area coverage is specified by the user in terms of  $\epsilon$  and  $\lambda$ , Figure 3 can be used to determine the number of sensors ( $M$ ) that should report data to the fusion center. Similar results were obtained for other values of the correlation coefficient  $\rho$ .

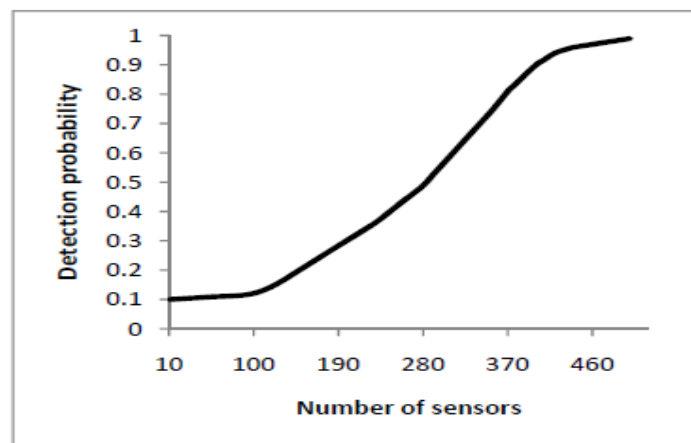


Figure 3. Variation of detection probability with the number of sensors for  $\epsilon=0.05$ ,  $P_0=100$  and  $n=2$ .

## 6.2. Performance Comparison of Sensor Selection Schemes

In this section, the sensor field is assumed to be densely deployed with  $N=500$  sensors and the fusion center selects a subset of  $M$  sensors to meet a specified coverage. For the target detection model, we set  $n=2$ ,  $P_0=100$  and assume that the noise variance is unity. We assume that each sensor quantizes its samples into a 6 bit data packet. We compare the performance of the proposed sensor selection schemes with CROSS and DRS described in [11].

Figures 4-6 show the performance comparison of different sensor selection schemes in terms of discrete joint entropy, for different values of the correlation coefficient  $\rho$ . The number of selected sensors ( $M$ ) correspond to a specific value of target detection probability  $\lambda$  for  $\varepsilon=5\%$ . The figures show that entropy based sensor selection scheme outperforms the other sensor selection methods for different values of  $\rho$ . The advantages of entropy based and distance based sensor selection scheme are evident for higher values of  $\rho$  as shown in Figures 4-6. As the sensor data becomes highly correlated, the entropy and distance based sensor selection schemes select the most informative subset of sensors significantly outperforming CROSS and DRS.

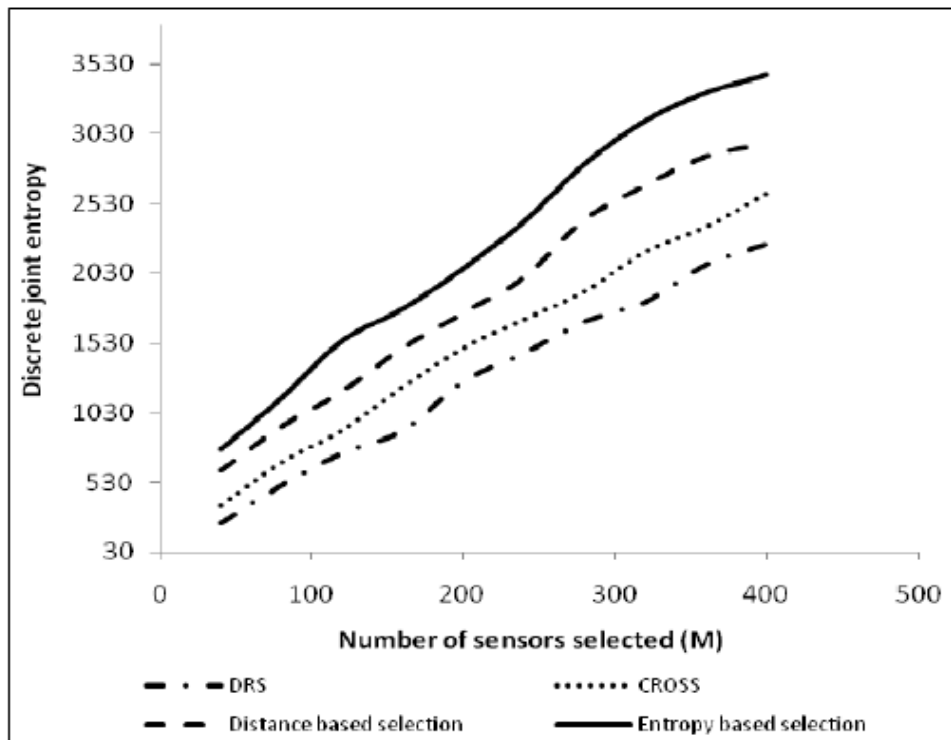


Figure 4. Performance comparison of sensor selection schemes for  $\rho=0.2$

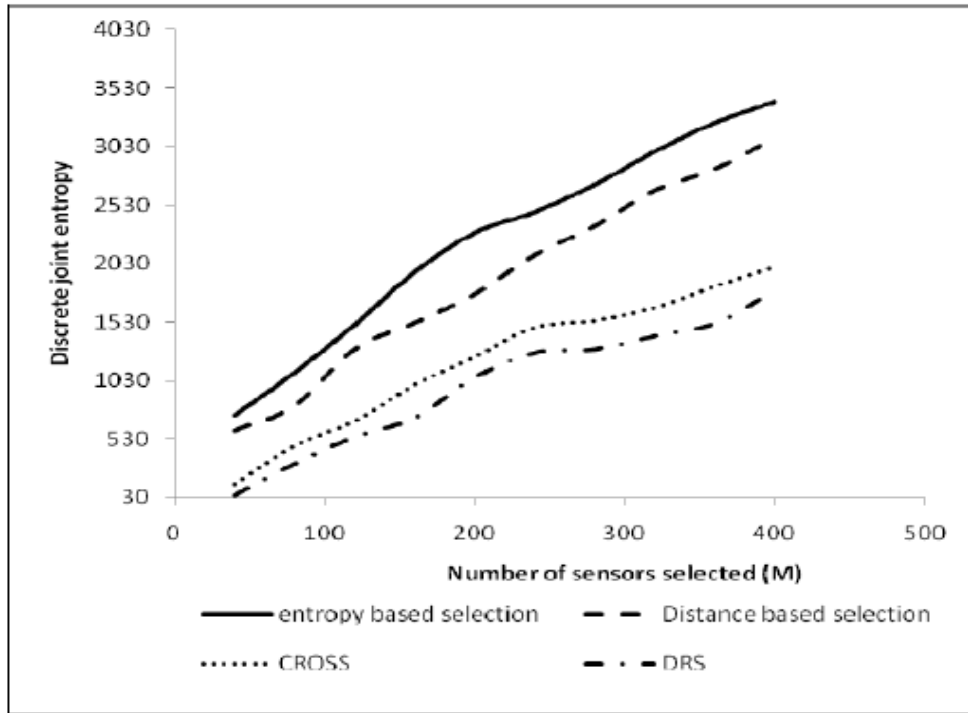


Figure 5. Performance comparison of sensor selection schemes for  $\rho=0.5$

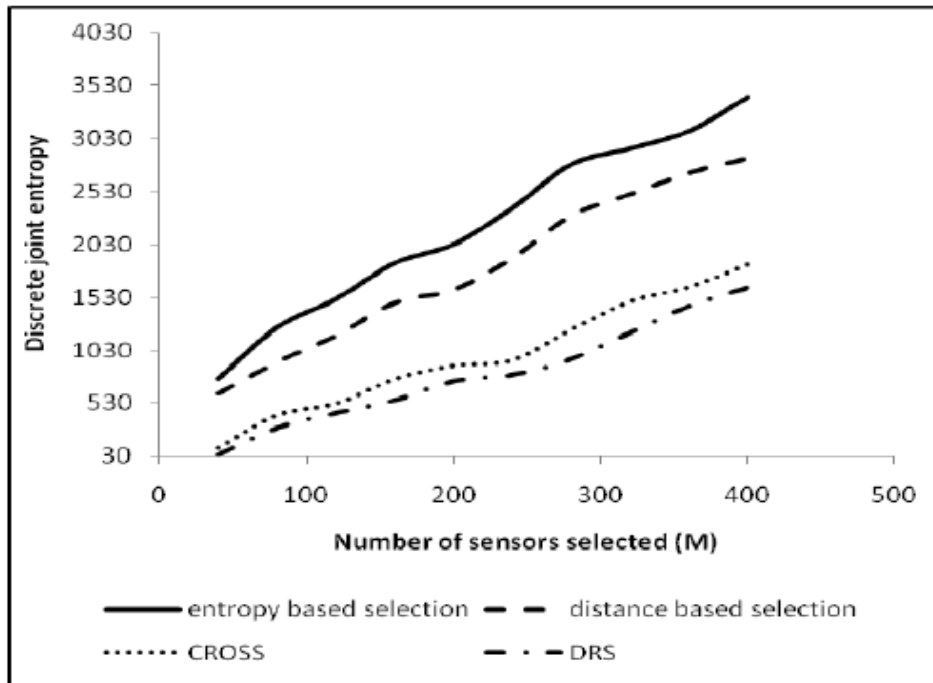


Figure 6. Performance comparison of sensor selection schemes for  $\rho=0.8$

The above results clearly indicate that the entropy based sensor selection scheme performs the best for various values of  $M$  and  $\rho$ . The distance based selection performs a little worse than the entropy based scheme but much better than DRS and CROSS.

Table I shows the percent improvement in discrete joint entropy of the entropy based sensor selection scheme when compared to CROSS for different values of  $M$  and  $\rho$ . We use equation (24) to compute the discrete joint entropy of the data collected by the selected sensors. We then evaluate the percent improvement in the discrete joint entropy. The results show that for a given  $M$ , the percent improvement increases with increase in  $\rho$ . This shows that for heavily correlated data, entropy based sensor selection offers a significant advantage.

Table 1. Percent improvement in discrete joint entropy: comparison of entropy based sensor selection and CROSS.

Number of sensors selected ( $M$ )	Percent improvement in discrete joint entropy		
	$\rho=0.2$	$\rho=0.5$	$\rho=0.8$
120	41	54	57
200	27	33	38
320	32	38	42
400	24	30	40

These results show that the entropy based sensor selection scheme achieves the best performance for different values of  $M$  (area coverage) and  $\rho$ . The distance based selection performs a little worse than the entropy based sensor selection. However, both the distance based and entropy based sensor selection schemes outperform CROSS and DRS.

The new schemes incur additional computational effort. For an  $M$  value of 200, the computational times required by DRS and CROSS were 3 seconds per communication round, on an Intel Pentium 4 processor (3.2 GHz, 2 GB RAM). The distance based sensor selection requires on an average 5 seconds per communication round while the entropy based sensor selection requires a relatively higher computational time of 8 seconds per communication round. Simulation results showed that the computational effort comparisons for different schemes were similar for other values of  $M$  and  $\rho$ . However, the advantages of the improvement in discrete joint entropy showed in Table I outweigh the additional computational cost.

## 7. CONCLUSIONS

This paper has proposed a new model for quantifying the area coverage in terms of the target detection and false alarm probabilities. Analytical solutions were developed for the distributed detection problem in the presence of correlated sensor noise. The mathematical analysis of the target detection and false alarm probabilities at the fusion center were challenging since the sensor decisions are correlated. However, assuming that the number of sensors is large, the central limit theorem was used to derive expressions for the global target detection and false

alarm probabilities. The results derived showed that the target detection probability, false alarm rate, and the area coverage were functions of the correlation coefficient.

In the second part of the paper, the sensor selection problem was formulated and solved using information theoretic measures such as discrete joint entropy. Two new sensor selection schemes were proposed for selecting the minimum number of sensors that achieve the desired target detection probability, selecting sensors which are as far apart from each other as possible. Extensive simulations were performed to compare different sensor selections schemes for various values of  $M$  and  $\rho$ . Simulation results showed that the entropy based and the distance based sensor selection schemes outperform CROSS and DRS. The performance improvement of the entropy based sensor selection was more evident for higher values of the correlation coefficient. Future work will examine the utility of other information theoretic measures such as K-L distance and mutual information for sensor selection problems.

## REFERENCES

- [1] S. Tilak, N.B. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensor network models," *ACM Mobile Computing and Communications Review*, vol.6, pp. 28-36, 2002.
- [2] D. Tian, and N. D. Georganas, "A coverage-preserving node scheduling scheme for large wireless sensor networks," *Proc. 1st ACM Workshop on Wireless Sensor networks and Applications*, 2002.
- [3] S. Shakkottai, R. Srikant, and N. Shroff, "Unreliable sensor grids: coverage, connectivity and diameter," *Proc. INFOCOM*, pp. 1073-1083, 2003.
- [4] W. Choi and S. K. Das, "Trade-off Between Coverage and Data Reporting Latency for Energy-Conserving data gathering in wireless sensor networks," *Proc. International conference on mobile ad hoc and sensor systems*, 2004.
- [5] S. Slijepcevic, and M. Potkonjak, "Power efficient organization of wireless sensor networks", *Proc. IEEE International conference on Communications*, pp. 472-476, 2001.
- [6] M. Cardei, D. MacCallum, X. Cheng, M. Min, X. Jia, D. Li, and D. Z. Du, "Wireless sensor networks with energy efficient organization," *Journal of Interconnection Networks*, vol. 3, no. 3-4, pp. 213-229, 2002.
- [7] M. Cardei, and D.Z. Du, "Improving wireless sensor network lifetime through power aware organization," *ACM Wireless Networks*, 11(3):333-340, 2005.
- [8] B. Liu, and D. Towsley, "A study of the coverage of large-scale sensor networks," *Proc. first IEEE International conference on mobile ad hoc and sensor systems*, Oct. 2004.
- [9] M. Cardei and J. Wu, "Energy efficient coverage problems in wireless ad hoc sensor networks," *Computer Communications*, vol. 29, no. 4, pp. 413-420, 2006.
- [10] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based Sensor Selection Heuristic for Localization," *ACM-IEEE third international symposium on information processing in sensor networks (IPSN)*, Berkeley CA, April 26-27, 2004.
- [11] W. Choi and S.K. Das, CROSS: "A probabilistic constrained random sensor selection scheme in wireless sensor networks," *Performance Evaluation*, vol.66, no. 12, pp. 754-772, 2009.
- [12] L. Doherty and K.S.J. Pister, "Scattered data selection for dense sensor networks," *Proc. Third international symposium on information processing in sensor networks*, pp. 369-378, 2004.
- [13] V. Aalo and R. Viswanathan, "Asymptotic performance of a distributed detection system in Gaussian noise," *IEEE Trans. on Signal Processing*, vol. 40, no.1, Jan. 1992.



- [14] Y. Zhang, M. Ramkumar, and N. Memon, "Information flow based routing algorithms for wireless sensor networks," *Proc. IEEE Global Telecommunications Conf. (GLOBECOM'04)*, vol. 2, pp. 742-747, Nov. 2004.
- [15] J. F. Chamberland and V. V. Veeravalli. "Decentralized detection in wireless sensor systems with dependent observations," *Proc. International conference on computing, communications and control technologies*, Aug. 2004.
- [16] R. Niu, P.K. Varshney, M. H. Moore and D. Kalmer "Decision fusion in a wireless sensor network with a large number of sensors," *Proc. 7th IEEE International Conference on Information Fusion (ICIF '04)*, Stockholm, Sweden, June–July 2004.
- [17] R. Niu and P.K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on wireless communications and networking*, vol. 4, 2005, pp. 462-472.
- [18] S. Patten, B. Krishnamachari and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," *ACM Trans. on sensor networks*, vol. 4, no.4, Aug. 2008.
- [19] S. Coleri and P. Varaiya, "Fault tolerance and energy efficiency of data aggregation schemes for sensor networks," *Proc. IEEE Vehicular Technology Conference*, Sept. 2004.
- [20] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. Wiley 1991.
- [21] D. Ganesan, R. Cristescu, and B. Beferull-Lozano, "Power efficient sensor placement and transmission structure for data gathering under distortion constraints," *Proc. third international symposium on information processing in sensor networks*, April 2004.
- [22] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of the approximations for maximizing submodular set functions," *Mathematical Programming* , vol. 14, pp. 265-294, 1978.