Distributed Signal Processing with Randomized Data Selection Based on the Generalized Approach in Wireless Sensor Networks

VYACHESLAV TUZLUKOV, WON-SIK YOON, and YONG DEAK KIM
Department of Electrical and Computer Engineering
College of Information Technology, Ajou University
San 5, Wonchon-dong, Paldal-gu, Suwon 442-749
KOREA, REPUBLIC OF
Email: tuzlukov@ajou.ac.kr, wsyoon@ajou.ac.kr, yongdkim@ajou.ac.kr

bstract: - Performing robust detection with resource limitations such as low-power requirements or limited ommunication bandwidth is becoming increasingly important in contexts involving distributed signal processng based on the generalized approach. One way to address these constraints consists of reducing the amount of ata used by the detection algorithms. Intelligent data selection in detection can be highly dependent on a priori nformation about the signal and noise. In this paper, we discuss detection strategies based on the generalized pproach to distributed signal processing with randomized data selection and analyze the resulting algorithms' erformance. Randomized data selection is a viable approach in the absence of reliable and detailed a priori inormation, and it provides a reasonable lower bound on signal processing performance as more a priori infornation is incorporated. The randomized selection procedure has the added benefits of simple implementation in distributed environment and limited communication overhead. We analyze a binary hypothesis testing probem using the generalized approach to signal processing in the presence of noise, and determine several useful roperties of the generalized detector derived from the decision-making rule. We show that the use of the generalized approach to signal processing in noise allows us to obtain the better wireless sensor performance in conparison with the systems using the Neyman-Pearson and matched filter detection algorithms. The advantages nd disadvantages of the of the generalized approach to signal processing with randomized data selection based n distributed sensor networks are also discussed.

ey-Words: - Distributed signal processing, generalized detector, randomized algorithms, randomized sampling, sensor networks

1 Introduction

With recent advances in device and computing technologies, the wireless sensor networks are becoming ncreasingly ubiquitous. Designing signal processing lgorithms that satisfy constraints imposed by these etworks is, therefore, becoming a necessity [1,2]. Specifically, these algorithms need to be efficient nd robust, and in the case of battery-powered wireess sensor networks, they need to operate under power constraints [3, 4]. In addition, there is usually the dded requirement of restricted communication bandwidth for wireless sensor networks and, therefore, careful management of the wireless sensor network's data transmission volume is important [5].

While it may be appropriate to design wireless sensor networks that densely populate a region with micro-sensors during sensor deployment, operation of the wireless sensor network may not require that all wireless sensor network nodes be operating and communicating at once. Indeed, for efficient operation, extended wireless sensor network lifetime, and efficient use of communication bandwidth, it may be desirable to select a subset of nodes to communicate at any fixed time [6]. The selected subset changes over time, varying usage among the nodes to extend their effective lifetime.

An appropriate signal processing algorithm based on the generalized approach to signal processing in noise [7—11] for node subset selection in a densely populated wireless sensor network can be highly dependent on the *a priori* information about the characteristics of both the signal and noise for a specific task or environment, and, consequently, it would be unreasonable to attempt to formulate a generic optimal procedure. In this paper, we explore the generalized approach to distributed signal processing with randomized data (node) selection in wireless sensor networks. Specifically, we abstract and simplify the problem by considering nodes as only communicating data for a signal processing problem. We use the average rate at which an individual node is included in the selected subset as the basic design parameter.

The selection reduces communication bandwidth requirements by limiting the amount of data transferred between nodes. Under the assumption that communication dominates the energy usage of the nodes, the wireless sensor network lifetime increases as this average rate decreases. Using these statements, we consider randomized data selection and analyze the wireless sensor network performance as a function of this average rate for detection algorithm based on the generalized approach to signal processing in noise.

Distributed wireless sensor networks are composed of interacting hardware and software systems. The energy efficiency of the wireless sensor networks can be improved by modifying the hardware [12] or any of the algorithms, such as data routing [13], source coding [14,15], or signal processing. In this paper, we focus on data selection based on the generalized approach to signal processing in noise as an algorithmic approach to improving the wireless sensor network's energy efficiency. Similar strategies have been used as a design for efficient systems in such diverse areas as filter approximation [16], statistical regression [17], and multiple-input multiple-output wireless communication [18,19]. Selection, by reducing the amount of communication congestion throughout the wireless sensor network and avoiding the computational burden of signal processing all available data, relieves two major sources of energy dissipation. Since we are concerned primarily with the selection algorithm and its impact upon signal processing performance, we do not attempt to quantify the energy savings because it would be highly dependent upon specific algorithms or hardware properties.

While data selection algorithms accounting for many aspects of the wireless sensor network's state can be useful in practice, we choose a generic approach requiring limited *a priori* information and communication overhead. Specifically, we consider a randomized data selection strategy. This approach leads us to useful signal detection algorithms based on the generalized approach to signal processing in noise in distinct areas such as estimation, hardware failure modeling, low-power design [20], and theoreticcal computer science [21].

2 Selection of Data

Our analysis focuses on signal processing procedures using data collected in a single time slot. We assume that each measurement is assigned an identifier index i, arbitrary chosen between 1 and N, where N is the total number of networks nodes. Subject to this model, the total data available in time slot m is denoted by

$$\mathbf{X}(m) = \begin{vmatrix} x_1(m) \\ x_2(m) \\ \vdots \\ x_N(m) \end{vmatrix} . \tag{1}$$

Under randomized selection rule, the decision to select measurement $x_a(m)$ depends on the outcome of an indicator random variable denoted $q_a(m)$. The random variable is independent of all other indicator random variables and from other physically measurable quantities available to the detector. In our model, each measurements in the current time slot is selected with the probability p_q, i.e., $q_a(m)$ has the probability mass function

$$f_{q_a(m)}(q) = \begin{cases} p_q\,, & q=1 \\ (1-p_q)\,, & q=0\,. \end{cases} \tag{2}$$

This selection rule reduces the expected complexity of the detector implementation by a factor of p_q.

The randomly selected data vector, $\mathbf{X}_q(m)$, can be represented by the equation

$$\mathbf{X}_q(m) = \mathbf{Q}(m) \cdot \mathbf{X}(m) \ . \tag{3}$$

Here, $\mathbf{X}(m)$ is the data vector defined in (1), and $\mathbf{Q}(m)$ is a $N \times N$ diagonal matrix with the i-th entry given by $\mathbf{Q}_{ii}(m) = q_i(m)$. Consequently, the vector $\mathbf{X}_q(m)$ is N-dimensional with each entry being either zero or a data measurement. In each time slot m, the number of nonzero entries of $\mathbf{X}_q(m)$ is a random variable, which we denote by $K(m)$. Prior to discussing the specific detection problems, we examine the signal statistics for $\mathbf{X}_q(m)$. Since the generalized detector only receivers a portion of the data, we must base our signal processing algorithm based on the generalized approach to signal processing upon the conditional density for $\mathbf{X}_q(m)$ given $\mathbf{Q}(m) = \mathbf{Q}$. To establish the notation for the conditional density, let the set

$$S(m) = \{j \mid \mathbf{Q}_{jj}(m) = 1\} \tag{4}$$

denote the selected measurements in time slot m. If a particular realization of this set is $S = \{j_1, j_2, \ldots, j_K\}$, the conditional density for $\mathbf{X}_q(m)$ is the joint density for $x_{j_1}(m), x_{j_2}(m), \ldots, x_{j_K}(m)$. The indicator random variables are independent of $\mathbf{X}(m)$, so no useful information about the signal is gained by observing

Q(m). For notational convenience, we will denote the conditional density of $\mathbf{X}_q(m)$ given $\mathbf{Q}(m)=\mathbf{Q}$ by the expression

$$f_{\mathbf{X}_q|\mathbf{Q}}(\mathbf{X}\,|\,\mathbf{Q}) \tag{5}$$

with the dependence on m understood.

3 Distributed Signal Processing

The properties of a wireless sensor network strongly influence the choice of appropriate distribute signal processing algorithms based on the generalized approach to signal processing in noise. Important variables influencing this choice include the number and density of nodes, the area covered by the wireless sensor network, available information about extremal environment, and the communication capabilities of the wireless sensor network.

Randomized data selection may be an attractive approach to distributed signal processing based on the generalized approach to signal processing in a variety of situations. For example, many simple nodes may be densely distributed throughout the extent of the wireless sensor network. Additionally, no *a priori* information about where targets are likely to appear may be available to aid the sensor selection algorithm. Due to the high node density, data from neighboring nodes can be highly redundant. In this situation, random selection with a small value of p_q can lead us to acceptable detector performance, while limiting the energy dissipated by communication of sensor data through the wireless sensor network. Additionally, the randomized selection procedure avoids computational or communication overhead that may be incurred from more complicated iterative selection procedures [22], or from centralized coordination of sensor selection.

Randomized selection is compatible with common architectures for *ad hoc* wireless sensor networking. Many wireless sensor networks use a combination local clustering(where a group of nodes communicate with a sink) and multihop routing in their wireless sensor networking protocols. Clustering can be combined with randomized data selection; in every time slot, each node in the cluster randomly determines whether to transmit its sensor measurement to the sink. The selection procedure limits the expected amount of data processed by each local sink. In multihop routing, each of selected sensor measurements follows a path through several nodes in the wireless sensor network. Any node in the wireless sensor network sees a random num routing, each of selected sensor measurements follows a path through several nodes in the wireless sensor network. Any node in the wireless sensor network sees a random number of packets from an individual time slot. Thus, if necessary, any node in the wireless sensor network can use the signal detection algorithms based on the generalized approach to signal processing in noise. Finally, note that the techniques we use to adapt the generalized detector to fluctuations in the size of $\mathbf{X}_q(m)$ can be applied to situations where unreliable sensor or communication hardware lead to intermittent loss of data in the wireless sensor network.

4 Detection of Signals

Let us analyze the interaction of randomized selection and signal detection in a background of additive Gaussian noise. The detector based on the generalized approach to signal processing in noise[7–11] from a binary hypothesis testing model is the likelihood ratio test. Test compares the likelihood ratio $L(\mathbf{X})$, defined as the ratio between the conditional probability distribution densities for $\mathbf{X}(m)$,with a fixed threshold K_g. If the generalized detector observes $\mathbf{X}$ in a region of sample space where $L(\mathbf{X})\geq K_g$, it makes the decision "a yes" signal — $\hat{H}=H_1$. Otherwise, it decides "a no" signal — $\hat{H}=H_0$. We denote decision rules of this form with the notation

$$L(\mathbf{X})\geq K_g \;\Rightarrow\; H_1 \text{ and } L(\mathbf{X})<K_g \;\Rightarrow\; H_0. \tag{6}$$

Our analysis of the likelihood ratio test highlights two key issues inherent to randomized data selection. First, we discuss the binary hypothesis test, and account for random selection in its statistical model. Second, we suggest low-complexity detectors based on the generalized approach to signal processing in noise that adapt to fluctuation in the amount of selected data. Additionally, we consider the robustness of the generalized detector to inaccuracies or unknown parameters in the *a priori* model for target signal's probability distribution density. While this issue is not directly related to random sampling, it illustrates the challenging signal processing environment in which the generalized detectors often operate. In order to derive useful properties of the likelihood ratio test in the presence of random selection, we impose restrictions on the statistical model for the target signature. To balance the generality of the signal model with its special statistical structure, we assume that the probability distribution density of the target signal is symmetric about the origin of the sample space. We shall refer to random vectors that satisfy this condition as even random vectors or even-symmetric signals.

This signal model establishes a useful structure on the probability distribution density of the signal, enabling us to determine key properties of the likelihood ratio test. Additionally, the signal model is broad enough to model many interesting target signatures. For example, a sinusoid with a unknown, uniformly distributed phase satisfies the condition as does a zero-mean, Gaussian random vector with a known covariance matrix in the following definition: *an N-dimensional random vector* $\mathbf{S}(m)$ *is referred to as even if, for every* $\mathbf{S}_0 \in \Re^N$, *its probability distribution density function satisfies the condition* $f_{\mathbf{S}}(\mathbf{S}_0) = f_{\mathbf{S}}(-\mathbf{S}_0)$. The general binary hypothesis test for signals in the additive Gaussian noise obeys the following statistical model:

$$H_0 : \begin{cases} \mathbf{X}(m) = \mathbf{n}(m), \\ \eta(m) = \mathbf{n}_1(m), \end{cases} \tag{7}$$

and

$$H_1 : \begin{cases} \mathbf{X}(m) = \mathbf{S}(m) + \mathbf{n}(m), \\ \eta(m) = \mathbf{n}_1(m), \end{cases} \tag{8}$$

where $\mathbf{n}(m)$ and $\mathbf{n}_1(m)$ are *N*-dimensional, zero-mean, Gaussian random vectors with the variance σ_n^2. The vector $\mathbf{n}_1(m)$ is the additional (reference) and uncorrelated with the vector $\mathbf{n}(m)$ noise. It is *a priori* known "a no" signal in the reference vector $\mathbf{n}_1(m)$ [7–11]. The signal vector $\mathbf{S}(m)$ has an even-symmetric probability distribution density. Finally, we assume that $\mathbf{S}(m)$, $\mathbf{n}(m)$, and $\mathbf{n}_1(m)$ are independent random vectors. This model describes the statistics of the data without randomized selection.

In the presence of randomized data selection, the generalized detector has access to the indicator random variables in $\mathbf{Q}(m)$ and processes the subset of the available data contained in $\mathbf{X}_q(m)$. The likelihood ratio for the generalized detector with randomized selection can be expressed as

$$L(\mathbf{X}_q, \mathbf{Q}) = \frac{f_{\mathbf{X}_q,\mathbf{Q}|H_1}(\mathbf{X}_q, \mathbf{Q} \mid H_1)}{f_{\eta,\mathbf{Q}|H_0}(\eta, \mathbf{Q} \mid H_0)}$$

$$= \frac{f_{\mathbf{X}_q|\mathbf{Q},H_1}(\mathbf{X}_q \mid \mathbf{Q}, H_1) \cdot f_{\mathbf{Q}|H_1}(\mathbf{Q} \mid H_1)}{f_{\eta|\mathbf{Q},H_0}(\eta \mid \mathbf{Q}, H_0)} = L(\mathbf{X}_q \mid \mathbf{Q}) \tag{9}$$

The simplification in the likelihood occurs because the indicator random variables are independent of the hypotheses H_i.

Since conditioning upon $\mathbf{Q}(m)$ does not affect the selected data in $\mathbf{X}_q(m)$, the detection problem based upon $\mathbf{X}_q(m)$ and $\mathbf{Q}(m)$ reduces to an unconditional detection problem for the data associated with the nonzero indicator random variables. For example, if three pieces of data are available, there are eight possible arrangements of the indicator random variables. If measurements 1 and 2 are selected in time slot *m*, the generalized detector must take a decision $\hat{H}$ based upon the joint probability distribution densities

$$f_{\mathbf{X}_q|\mathbf{Q},H}(\mathbf{X} \mid 1,2,H_0) = f_{x_1,x_2|H}(x_1, x_2 \mid H_0) \tag{10}$$

and

$$f_{\mathbf{X}_q|\mathbf{Q},H}(\mathbf{X} \mid 1,2,H_1) = f_{x_1,x_2|H}(x_1, x_2 \mid H_1) \ . \tag{11}$$

Likewise, if measurements 2 and 3 are selected, the decision $\hat{H}$ is determined from $f_{x_2,x_3|H}(x_2, x_3 \mid H_0)$ and $f_{x_2,x_3|H}(x_2, x_3 \mid H_1)$.

Based upon (9), the likelihood ratio test for $\mathbf{X}_q(m)$ and $\mathbf{Q}(m)$ reduces to the comparison of $L(\mathbf{X}_q \mid \mathbf{Q})$ to a fixed threshold. Under the use of optimal detectors based on classical and modern signal detection theories, while the test is optimal under the Neyman-Pearson detection criteria, it poses some practical problems. First, determining the threshold can become computationally complex when there is a large amount of data available for selection. Under the use of the detector based on the generalized approach to signal processing in noise, determining the threshold is independent of a large amount of data available from selection. In this case, the threshold is defined by the background noise of the generalized detector, and is not computationally complex. The threshold that achieves a desired false alarm rate P_F under the use of the generalized detector is determined by inverting the equation

$$P_F(K_g) = \sum_{\mathbf{Q}} \Pr[L(\mathbf{X}_q) > K_g \mid \mathbf{Q}, H_0] \ . \tag{12}$$

If *N* samples of data are available, there are 2^N terms in the summation. The functions of the threshold K_g given by (12) may be easily parameterized in comparison with the optimal detectors based on classical and modern signal detection theories. Second, since the threshold K_g is constant while $\mathbf{Q}$ fluctuates, under the use of detectors based on classical and modern signal processing approaches, the conditional false alarm rate

$$P_F(\mathbf{Q},K_g)=\Pr[L(\mathbf{X}_q)>K_g\,|\,\mathbf{Q},H_0] \quad (13)$$

fluctuates as well. In a situation where actions taken following a false alarm are costly, however, this fluctuation may not be desirable, since it is induced by the random data selection rather than an information-bearing signal. Under the use of the generalized detector based on the generalized approach to signal processing in noise, the conditional false alarm rate is fixed both for each realization $\mathbf{Q}$ and for all available data. Therefore, we have not any constraints in the conditional false alarm rate under the use of the generalized detector in the case of distribute signal processing with randomized data selection in wireless sensor networks.

5 Detection of Sinusoidal Signal

Consider an example – detecting a sinusoidal signal. We consider detection of a sinusoidal signal in the presence of randomized data selection. Our analysis illustrates the difficulties associated with detection in the presence of uncertainty in the target signal. Consider a set of data generated by sampling a signal at several locations, denoted by υ_i, $i=1,\ldots,N$. We shall assume that these locations can be modeled by a set of independent, identically distributed uniform random variables over an interval significantly larger in comparison with the sinusoid's wavelength.

Let the hypothesis H_0 denote the state in which the sinusoid is absent, and the hypothesis H_1 denote the state when it is present. The i–th measurement under each hypothesis is given by

$$H_0:\begin{cases} x_i=n_i\,, \\ \eta_i=n_{1_i} \end{cases} \quad (14)$$

and

$$H_1:\begin{cases} x_i=A\cos\left(\dfrac{2\pi\upsilon_i}{\lambda}+\varphi\right)+n_i\,, \\ \eta_i=n_{1_i}\,. \end{cases} \quad (15)$$

The random variables n_i and n_{1_i} are the zero-mean Gaussian random variables with the variance σ_n^2. The probability distribution density for $\mathbf{X}_q(m)$, conditioned upon $\mathbf{Q}(m)$ and the hypothesis H_0 is defined by the Bessel function of the second order of imaginary argument under the use of the generalized detector based on the generalized approach to signal processing in noise:

$$f_{n_i^2-n^2}(z)=\frac{1}{2\pi\sigma_n^2}K_0\left(\frac{z}{2\sigma_n^2}\right) \quad \text{as} \quad N\to 0 \quad (16)$$

or

$$f_{n_i^2-n^2}(z)=\frac{1}{2\sigma_n^2}\sqrt{\frac{\rho}{\pi}}\cdot e^{-\frac{\rho z^2}{4\sigma_n^4}} \quad \text{as} \quad N\to\infty\,. \quad (17)$$

In order to determine the likelihood ratio and the resulting generalized receiver operating characteristics, we also need the probability distribution density for $\mathbf{X}_q(m)$, conditioned upon $\mathbf{Q}(m)$ and the hypothesis H_1. This conditional probability distribution density depends, in turn, on the joint probability distribution density of

$$w_i=\frac{2\pi\upsilon_i}{\lambda}+\varphi \quad (18)$$

for the selected data in $S(m)$. The probability distribution density for the signal is a function of the joint probability distribution density of the phase random variables. Since the signal and noise are independent under the hypothesis H_1, the overall conditional probability distribution density for $\mathbf{X}_q(m)$ is the convolution of the signal probability distribution density and the noise probability distribution density. The determination of the joint probability distribution density for the phase random variables is a key step in this calculation.

Since $\{\upsilon_i\}$ are independent and uniform over a large interval, we can approximate $\{w_i\}$ as independent, identically distributed uniform random variables over the region $[-\pi,\pi]$. Using this model, we can analyze the form of the likelihood ratio test for the model suggested in (14) and (15). Here, we assume that the sink knows the value of A exactly. The signal is $\mathbf{a}(m)$, where $\mathbf{a}(m)$ is a K-dimensional random vector. Each entry takes the form

$$a_i(m)=A\cos(w_i)\,. \quad (19)$$

Based upon our approximation, the probability distribution density for $\mathbf{a}(m)$ is determined in the following form

$$f_{\mathbf{a}|K}(\mathbf{a}\,|\,K)=\prod_{i=1}^{K}\frac{u(A-|a_i|)}{\pi\sqrt{A^2-a_i^2}}\,, \quad (20)$$

where $u(.)$ denotes the unit step function. This probability distribution density is nonzero over the K-dimensional hypercube of side A. For fixed K, we denote the randomly selected data by $\mathbf{X}_K(m)$. When applied to a vector, the subscript K indicates its dimension. This does not contradict our earlier notation, where the subscript of a scalar random variable indicated the identity of the measurement. The dimension subscript is always attached to a vector, not a scalar. This random vector lists the selected data contiguously, rather than with zeros as in $\mathbf{X}_q(m)$. For notational convenience, we assume that measurements 1 to K are selected, so $\mathbf{X}_K = [x_1, x_2, ..., x_K]^T$. This notation does not reduce the applicability of the analysis, since our modeling assumptions make the measurements statistically indistinguishable. Their joint statistics depend only on K and not on the measurement identifiers. The resulting signal model has the following form

$$H_0 : \begin{cases} \mathbf{X}_K(m) = \mathbf{n}(m), \\ \eta_K(m) = \mathbf{n}_1(m) \end{cases} \tag{21}$$

and

$$H_1 : \begin{cases} \mathbf{X}_K(m) = \mathbf{a}(m) + \mathbf{n}(m), \\ \eta_K(m) = \mathbf{n}_1(m). \end{cases} \tag{22}$$

Based on these probability distribution density functions, we can construct the likelihood ratio test for fixed values of K and A. The conditional probability distribution density under the hypothesis H_0 is zero mean and determined by (16) or (17). Under the hypothesis H_1, the conditional probability distribution density is the convolution of (16) or (17) with the probability distribution density for $\mathbf{a}(m)$ given in (20). The conditional probability distribution density for $\mathbf{X}_K(m)$ under the hypothesis H_1 can be written in terms of a one-dimensional $(1-D)$ convolution, since both conditional probability distribution densities are separable. The conditional probability distribution density takes the following form

$$f_{\mathbf{X}_K|K,H_1}(\mathbf{X}|K,H_1) = f_{\mathbf{a}|K}(\mathbf{X}|K) * f_{n_1^2-n^2|K}(\mathbf{X}|K)$$

$$= \prod_{i=1}^{K} \int_{-\infty}^{\infty} \frac{u(A-|a_i|)}{\pi\sqrt{A^2-a_i^2}} \frac{1}{2\pi\sigma_n^2} K_0\left[\frac{(z_i-a_i)^2}{2\sigma_n^2}\right]$$

$$\text{as} \quad N \to 0 \tag{23}$$

or

$$f_{\mathbf{X}_K|K,H_1}(\mathbf{X}|K,H_1) = f_{\mathbf{a}|K}(\mathbf{X}|K) * f_{n_1^2-n^2|K}(\mathbf{X}|K)$$

$$= \prod_{i=1}^{K} \int_{-\infty}^{\infty} \frac{u(A-|a_i|)}{\pi\sqrt{A^2-a_i^2}} \frac{1}{2\sigma_n^2} \sqrt{\frac{\rho}{\pi}} \cdot e^{-\frac{\rho(z_i-a_i)^2}{4\sigma_n^4}} .$$

$$\text{as} \quad N \to \infty \tag{24}$$

Substituting (23) and (24) in terms of (16) and (17) in (9), we can obtain the likelihood ratio test for the generalized detector based on the generalized approach to signal processing in noise under the distributed signal processing with randomized data selection in wireless sensor networks.

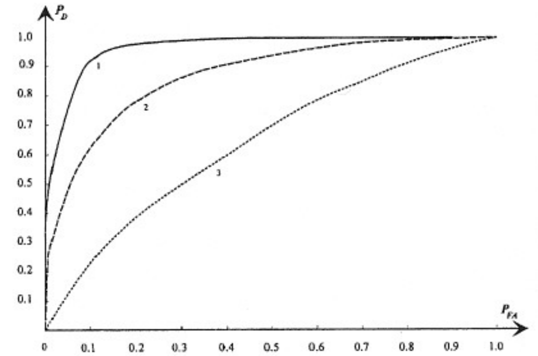

Fig.1. Detection performance comparison: 1 – generalized detector; 2 – Neyman-Pearson detector; 3 – matched filter.

Typically, the performance of the generalized detector is shown by operating characteristic, which plots the detection probability P_D as a function of the false alarm probability P_F (see Fig. 1, the number of Monte Carlo realizations used was 10 000). Reference to Fig. 1 shows us a great superiority under the use of the generalized detector in distributed signal processing under randomized data selection in wireless sensor networks in comparison with the matched filter and Neyman-Pearson detection algorithms.Both P_D and P_F can be calculated by integrating, respectively, the conditional probability distribution densities $f_{\mathbf{X}_K|K,H_1}(\mathbf{X}|K,H_1)$ and $f_{\mathbf{X}_K|K,H_0}(\mathbf{X}|K,H_0)$ over the $\hat{H} = H_1$ decision region. Thus, the operating characteristic of the generalized detector is generated as the threshold ranges over $0 \le K_g < \infty$. It can be shown that the operating characteristic of the generalized detector calculated from the likelihood ratio test gives the maximum achievable P_D for each false alarm rate $0 \le P_F \le 1$.

For $K = 1$, the generalized detector has an important universality property over the set of binary hypothesis tests for $A > 0$. The threshold K_g that achieves a certain P_F can be determined in terms of the Q function and $4\sigma_n^4$. Since the threshold K_g can be determined without knowledge of the wave amplitude A, the likelihood ratio test for the generalized detector based on the generalized approach to signal processing in noise under the distributed signal processsing with the randomized data selection in wireless sensor networks is a uniformly most powerful test. For such test, the decision regions that maximize P_D subject to a constraint on P_F are invariant to the actual value of the parameter A. The actual value of P_D, however, does not depend on A.

The contrast between the likelihood ratio test for $K=1$ and $K=2$ indicates some implementation challenges in the presence of uncertain signal models and random data selection. When $K>1$, the likelihood ratio test for $\mathbf{X}_K(m)$ is not a function of the received data magnitude. Since the likelihood ratio is increasing in all directions, the likelihood ratio test will declare $\hat{H}=H_0$ in a simply connected region containing the origin. Outside this region, it will declare $\hat{H}=H_1$. Thus, the two-dimensional $(2-D)$ test determines a closed curve that gives the boundary between the decision regions for $\hat{H}=H_0$ and $\hat{H}=H_1$. The implementation of the likelihood ratio test for the generalized detector based on the generalized approach to signal processing in noise under distributed signal processing with randomized data selection in wireless sensor networks is more complicated in two dimensions than in one. Finally, the likelihood ratio test's decision regions depend on the value of K. The shape of the decision regions varies as K changes. Evidently, the larger values of K lead to more complicated decision regions. For example, the decision regions for $K=2$ can be complicated sets in the (x_1, x_2) plane.

The difficulty in determining the decision regions under uncertainty in A and K makes the exact likelihood ratio test on $\mathbf{X}_K(m)$ challenging to implement. First, the fluctuation in K means that the generalized detector must be able to quickly adapt the decision regions for each time slot. Second, potential uncertainties in the target signal probability distribution density prevent the generalized detector from determining the exact likelihood ratio test. These challenges in the example detection problem persist for the general even signal model.

6 Conclusion

In this paper, we propose the generalized detector based on the generalized approach to signal processing in noise under distributed signal processing with randomized data selection in wireless sensor networks as a technique to cope with limited communication and computation resources in distributed wireless sensor networks. To illustrate the impact of randomized selection on signal processing by the generalized detector performed by wireless sensor network, we focus on a binary hypothesis testing problem: detection of a random vector with an even-symmetric probability distribution density in the Gaussian noise.

The detection problem shows several issues inherent in randomized data selection. First, understanding the impact of randomized data selection on the signal statistics is required to determine the new generalized detector structure. Our assumptions about the wireless sensor network's communication protocols allow us to design the generalized detector upon the conditional densities of the signals. Second, the fluctuation in the size of the selected sensor subset presents implementation challenges for the generalized detector. It does not need to perform the likelihood ratio test for a large number of potential subsets. Finally, we consider the impact of uncertainties in the *a priori* model for the signal probability distribution density on the generalized detector. This issue was independent of randomized data selection, but it describes the potential applications of the wireless sensor networks. They may often be used in situations where the coefficients of the signal probability distribution density are not known in advance.

The proposed likelihood ratio test for the generalized detector rests upon the statistical model for the noise. In practice, the noise is not Gaussian, as a rule. The generalized signal detection algorithm, however, serves as a template for designing generalized detectors for use with random sensor selection. Generalized detectors assuming different noise statistics will likely require similar approximations for the decision thresholds. Overall, the generalized signal processing algorithm serves as a low-complexity baseline for evaluating the performance of sensor selection and generalized signal detection algorithms in wireless sensor networks.

Future work on distributed signal processing and sensor selection can take many directions. The analysis of random sensor selection can be expanded to improve the performance baseline it provides. The selection algorithm can be modified to account for selection over multiple time slots or selection in the presence of colored noise. Both cases introduce an additional degree of freedom to the basic approach analyzed here. Additionally, exploring the interaction between selection and routing or distributed source coding algorithms may provide another interesting technique for balancing signal processing performance with communication cost.

Finally, the implementation of the generalized detector under distributed signal processing with the randomized data selection may be useful in practical employment of wireless sensor networks. The issues of robustness, complexity, and energy efficiency are significant in this new environment, and randomized selection provides a way to balance these performance criteria. The use of the generalized detector under distributed signal processing with the randomized data selection has several implications for real systems. First, it gives a performance baseline to algorithms based upon more detailed state information and statistical models. Second, it may provide a desirable operating point in the tradeoff between robustness, complexity, and performance. The generalized detector with random selection requires no extra communication to collect the wireless sensor network state, and is robust to a wide class of signal models. While it will likely produce performance inferior to signal processing algorithms based upon detailed models, it may be more robust to modeling errors, and may strike a desirable balance between performance and complexity. Third, the generalized detector with randomized data selection is compatible with the cluster-based distributed signal processing techniques proposed for wireless sensor networks. In wireless sensor networks where nodes cooperate in local clusters, the generalized detector with random data selection can be applied in each cluster. Finally, the generalized detector with random data selection performance depends only upon a single parameter. While we focused on single-shot processing in this paper, these signal detection algorithms based on the generalized approach to signal processing in noise will be used over time. The selection procedure must balance the information quality in an individual time slot with longevity of the wireless sensor network. Since the generalized detector with randomized data selection performance depends on a single parameter, it may be possible to illustrate the tradeoff between performance and wireless sensor network lifetime in appealing way through further analysis of this class of sensor selection algorithms.

Acknowledgment:
This work was supported in part by participation within the limits of the project "A Study on Wireless Sensor Networks for medical Information" sponsored by IITA, Korea.

References:

[1] S. Kumar, F. Zhao, and D. Shepherd, Collaborative Signal and Information Processing in Microsensor Networks, *IEEE Signal Processing Mag.*, Vol. 19, No. 3, 2002, pp. 13—14.

[2] J. Chen, K. Yao, and R. Hudson, R., Source Localization and Beamforming, *IEEE Signal Processing Mag.*, Vol. 19, No. 3, 2002, pp.30—39.

[3] D. Li, K. Wong, Y. Hu, and A. Sayeed, Detection, Classification, and Tracking of Targets, *IEEE Signal Processing Mag.*, Vol. 19, No. 3, 2002, pp. 17 —29.

[4] S. Blatt, Distributed Sensor Fusion for Sensor Networks, in *Proc. 4th Annu. Conf. Information Fusion*, 2001, pp. TuC3—TuC8.

[5] V. Raghunathan, C. Schrugers, S. Park, and M. Srivastava, Energy-Aware Wireless Microsensor Networks, *IEEE Signal Processing Mag.* Vol.19, No. 3, 2002, pp. 40—50.

[6] F. Zhao, J. Shin, and J. Reich, J., Information-Driven Dynamic Sensor Collaboration, *IEEE Signal Processing Mag.*, Vol. 19, No. 3, 2002, pp. 61—72.

[7] V. Tuzlukov, *Signal Processing in Noise: A New Methodology*, IEC, Minsk, 1998.

[8] V. Tuzlukov, A New Approach to Signal Detection, *Digital Signal Processing: A Review Journal*, Vol. 8, No. 3, 1998, pp. 166—184.

[9] V. Tuzlukov, *Signal Detection Theory*, Springer-Verlag, New York, 2001.

[10] V. Tuzlukov, *Signal Processing Noise*, CRC Press, Boca Raton, London, New York, Washington DC, 2002.

[11] V. Tuzlukov, V., *Signal and Image Processing in Navigational Systems*, CRC Press, Boca Raton, London, New York, Washington DC, 2004 (in press).

[12] A. Chandrakasan, R. Amirtharajah, S. Cho, J. Gooman, G. Konduri, J. Kulik, W. Rabiner, and A. Wang, Design Consideration for Distributed Microsensor Systems, in *Proc. CICC 1999*, pp. 279—286.

[13] D. Estrin, and R. Givindan, Next Century Challenges: Scalable Coordination in Sensor Networks, in *Proc. Mobicom 1999*, pp. 263 —286.

[14] S. Pradhan, J. Kusuma, and K. Ramchandran, Distributed Compression in a Dense Microsensor Network, *IEEE Signal Processing Mag.*, Vol. 19, No. 3, 2002, pp. 51—60.

[15] S. Draper, Successive Structuring of Source Coding Algorithms for Data Fusion, Buffering, and Distribution in Networks, Ph.D. Dissertation, M.I.T., Cambridge, MA, 2002.

[16] D. Mattera, F. Palmieri, and S. Haykin, Efficient Sparse FIR Filter Design, in *Proc. ICASSP 2002*, Vol. 2, pp. 1537—1540.

[17] A. Miller, *Subset Selection in Regression,* Chapman & Hall, New York, 1990.

[18] D. Gore, R. Nabar, and A. Paulraj, Selecting an Optimal Set of Transmit Antennas for a Low Rank Matrix Channel, in *Proc. ICASSP 2001,* pp. 2785—2788.

[19] A. Molisch, M. Win, and J. Winters, J., Capacity of MIMO Systems with Antenna Selection, in *Proc. ICC 2001,* pp. 570—574.

[20] M. Said, and A.Oppenheim, Discrete-Ttime Randomized Sampling, in *Proc. 8th Annu .IEEE Conf. Electronics, Circuits, and Systems,* Vol. 3, 2001, pp. 1407—1411.

[21] R. Motwani, and P. Raghavan, P., *Randomized Algorithms,* Cambridge Univ. Press, Cambrdge U.K., 1995.

[22] M. Nafie, A. Tewfik, and M. Ali, Deterministic and Iterative Solutions to Subset Selection Problems, *IEEE Trans.,* Vol. SP-50, No. 7, 2002, pp. 1591—1601.