Stochastic Binary Sensor Networks for Noisy Environments

T. Nguyen*
School of Electrical Engineering and Computer Science,
Oregon State University
E-mail: thinhq@eecs.orst.edu
*Corresponding author

D. Nguyen
School of Electrical Engineering and Computer Science,
Oregon State University,
E-mail: nguyendo@eecs.orst.edu

H. Liu
School of Electrical Engineering and Computer Science,
Oregon State University,
E-mail: hliu@eecs.orst.edu,

D. Tran
Department of Computer Science,
University of Dayton, Ohio,
E-mail: Duc.Tran@notes.udayton.edu

Abstract: This paper proposes a stochastic framework for detecting anomalies or gathering events of interest in a noisy environment using a network consisting of binary sensors. A binary sensor is an extremely coarse sensor, capable of measuring data to only 1-bit accuracy. Our proposed stochastic framework employs a large number of *inexpensive* binary sensors operating in a noisy environment, yet collaboratively they are able to obtain accurate measurements. The main contributions of this paper are: (a) The theoretical accuracy analysis of the proposed stochastic binary sensor network in noisy environments, (b) an adaptive data collection framework based on the current measurements in order to reduce the energy consumption, and (c) a novel coding scheme for energy-efficient routing. To quantify the performance of our proposed stochastic approach, we present the simulation results of two stochastic binary sensor networks for anomaly detection using our proposed coding scheme and adaptive data gathering framework. We demonstrate that our proposed framework can potentially reduce the energy consumption over the traditional approach by an order of magnitude.

1 Introduction

In recent years, sensor networks have emerged as an important class of networks for many military and commercial applications (1)(2)(3). A sensor network is a collection of wireless communication nodes. Each node is capable of sensing the environment and communicating the measured data to the neighboring nodes, and eventually to the external users. The majority of sensor networks are designed to collect data (4)(5) or to perform anomaly detection (6)(7). They are designed to achieve *accuracy*, *robustness*, and *energy efficiency*.

Because of the constraints on the energy consumption and/or the technologies, a sensor may be forced to reduce its sampling resolution either in time or in amplitude. This process results in lower accuracy of the measured data. Furthermore, the accuracy of the measured data can also be affected by the environmental noise. A special and important situation arises when all the sensors measure the same underlying signal x. However due to the environmental noise, each sensor i measures a different value $x_i = x + n_i$ where n_i is an additive noise sample. In this scenario, the objective of a sensor network is to accurately determine the underlying value x through collaboration among the sensors. For example, the nodes in a sensor network for anomaly or intrusion detection can exchange correlated data among each other to increase the detection accuracy. A higher correlation of measured data at different sensors leads to a stronger belief about the accuracy of the data.

Beside the accuracy issue, a sensor network must also be robust against sensor's failures since these sensors typically operate in an outdoor environment where they are subjected to harsh conditions. A straightforward approach to improve the robustness is to increase the number of sensors. However, it is desirable to construct a sensor network such that the accuracy of the collected data degrades gracefully in the presence of sensor failures.

Finally, since a sensor is a battery-operated device, minimizing energy consumption should also be considered. There have been many contributions in the area of protocol and system design for energy-efficient sensor networks (8)(9)(10)(11). If the measured data among the nodes are spatially correlated, a node can jointly compress its data and its neighbor's data in order to decrease the transmission energy (12)(13). A higher correlation of data measured among the sensors results in more energy saving.

In this paper, we present a stochastic binary sensor network that (a) achieves good accuracy, (b) enables the graceful degradation of data quality in the presence of sensor's failures, and (c) reduces the energy due to the adaptive data collection technique and a novel coding scheme. In the proposed approach, each sensor transmits only 1 bit of information per measurement, but collectively, accurate data measurements can be obtained. By allowing only 1-bit measurement, the sensors can be made at low cost and, therefore, are easy to replace or to discard. Furthermore, the proposed sensor network employs the stochastic approach which allows a graceful degradation of data quality (accuracy) in the presence of sensor failures. While we argue for robustness in terms of graceful degradation of data accuracy, the robustness of a sensor network ultimately depends on the routing algorithms to route the data around the failed sensors. Due to the limited scope of the paper, we do not discuss such routing algorithms. Instead, we refer the readers to many existing robust routing algorithms in the literature (5)(14)(15).

Our paper is organized as follows. In Section 2, we discuss some related work on sensor networks. In Section 3, we present our proposed model of stochastic binary sensor network. Since our model is not a typical sensor network model, we will motivate the use of such model with an intrusion detection application that employs inexpensive magnetic sensors. Section 4 is devoted to the accuracy analysis of our proposed stochastic binary sensor network in a variety of environments. In Section 5, we present a novel coding scheme to be used for energy-efficient routing. Section 6 provides the simulation results for two binary sensor networks that employ the adaptive data collection technique and the coding scheme to reduce energy consumption. Finally, we provide a few concluding remarks in Section 7.

2 Related Work

Since sensors are battery-operated devices, energy efficiency is one of the key considerations in designing a long-lived sensor network. As such, in recent years, there has been a vast literature on techniques for achieving good trade-offs between energy usage and the accuracy of the collected data. One approach to minimize energy consumption is the use of *in-network* processing. (16)(17)(18). This technique assumes that the measured data are spatially correlated, thus the sensors can jointly compress the data to result in fewer transmissions. For example, Cristescu *et al.* recently propose a simple model of the measured data based on the correlation coefficient. Using this model, the authors devise the data compression and routing algorithms in order to minimize the energy consumption in a network. (12)(17)(19). These algorithms aim to maximize data accuracy (or minimize the distortion) under the communication energy constraints. They are also designed with the assumption that each sensor is able code the data at a high resolution. On the other hand, our approach uses a large number of extreme coarse sensors, yet collaboratively these sensors can obtain high resolution data in noisy environments.

Other significant theoretical contributions on achieving trade-off between data accuracy and the number of bits used to represent data (i.e., energy consumption) is the work of Ishwar *et al.* on the principle of *bit conservation* (18). In this work, Ishwar *et al.* address the general problem of sampling bandlimited sensor fields in a distributed, limited-precision, communication-constrained, processing environment. In particular, *conservation of bits* principle states that "the bit budget per Nyquist-interval (the rate) can be distributed along the amplitude-axis (sensor precision) and space (sensor density) in an almost arbitrary discrete-valued manner, while retaining the same error-rate characteristics." The paper proposes methods that employ *interpolating function* to enable low precision sensors to achieve high resolution data when a large number of these sensors are used. Thus, our proposed binary stochastic sensor network is a special case of such formulation in which, the sensor precision is extremely coarse. On the other hand, Ishwar *et al.* focus exclusively on the data accuracy with respect to sensor precision and sensor density trade-offs without taking into consideration of the noisy measured data due to the environment. In our work, we incorporate the environmental noise, specifically the uniform and Gaussian noises, to derive the optimal estimators under such conditions. Luo also proposes similar method for estimating a corrupted signal due to random noise (20). In this method, each sensor collects one noise-corrupted sample, performs a local data quantization according to some probabilistic rule, and transmits the discrete messages to its neighbors. These discrete messages are dispersed through the network. Each sensor then uses these messages to compute its own minimum mean squared error (MMSE) estimate of the sample.

Other contributions on aggregation of correlated data in sensor networks have been advanced by Enachescu and Sharaf in (21)(22). In (21), Enachescu *et al.* propose a simple randomized algorithm for routing data on a grid of sensors in a way that promotes data aggregation. They show that their randomized algorithm is a constant-factor approximation to the optimal aggregation tree. This work mainly focus on energy efficient routing rather than achieving trade-off between the data accuracy and energy consumption of the network.

In terms of estimating the data under uncertainty, our work is probably most related to the recent work by Xiao *et al.* (23). In this work, the authors propose a distributed scheme to estimate the measured signal based on the average consensus. The objective of this distributed scheme is for every sensor to obtain the accurate measurements after exchanging data among each other for a number of rounds. On the other hand, our work focuses on data gathering from all the sensors to a processing node. Furthermore, unlike (23), our work makes use of coarse binary sensors, and provides the adaptive data gathering and coding schemes for reduction in energy consumption.

From the signal processing community, our work is most related to the work of Cvetkovic and Daubechies (24). This work provides an algorithm for achieving high resolution data by oversampling the data using only 1 bit resolution.

In terms of applications, our work is related to sensor networks for detecting the presence of intruders or chemical agents. In this scenario, a single sensor node often cannot detect the movement of intruders with high accuracy due to the noise induced in the measurement process or by the environment. However, if a sensor's neighbors also report similar results, the probability of having an intruder increases substantially. In (7), Vercauteren *et al.* propose to increase the accuracy of target tracking and classification using collaborative sensor networks. The literature also numbers several contributions in which, the sensors collaborate with each other to localize anomalies (6). In (6), Du *et al.* propose a number of metrics and methods for localizing the location attacks in sensor networks. The location attacks aim to alter the location information of the sensor nodes by sending out incorrect location information. Since each sensor node relies on the relative location of other sensors to compute its location, such an attack can potentially cause much damage in the entire network. Du *et al.* devise some heuristic algorithms for computing the difference between a node's actual location and the measured location based on other node's information. An attack is present if this difference is greater than a certain threshold. Similarly, our proposed sensor network also uses a threshold for determining anomalies. On the other hand, our work focuses on the trade-off between energy consumption and accuracy.

3 Network Model and Rationale

In this section, we discuss the mathematical model of data estimation in a stochastic binary sensor network and provide a specific motivation for such model.

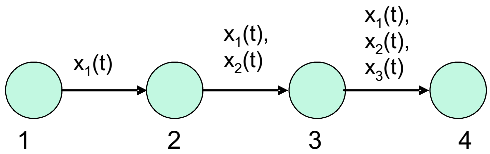


Figure 1: *A simple sensor network. Measured data are relayed to node 4.*

3.1 Mathematical Model

For the purpose of illustration, we use a simple network consisting of four nodes in a straight line as shown in Figure 1. We assume that the data $x(t)$ at time t is identical (or highly correlated) within the coverage of the sensors. In other words, the data may vary temporally but not spatially. Each node i is to measure the underlying data $x(t)$ at time t. Theoretically, without noise, all the nodes would obtain identical measurements at time t. However, in non-ideal situations where either internal or external noise is present, an accurate measurement cannot be obtained using only a single node. Instead, the measured data at each sensor at time t is

$$x_i(t) = x(t) + n_i(t), \tag{1}$$

where $n_i(t)$ is independent and identically distributed noise at node i. The measurements from different nodes are forwarded to a processing node whose task is to accurately determine the value $x(t)$. For example, if node 4 is chosen to be the processing node, then it is responsible for determining the value of $x(t)$ based on all the measurements it receives from nodes 1, 2, and 3. Figure 1 shows how measurements are forwarded from nodes 1, 2, and 3 to node 4.

A popular method for estimating value of $x(t)$ (25) is given by

$$\hat{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t). \tag{2}$$

where N is the number of sensors. The larger the value of N, the higher accuracy of the estimate.

This method assumes an accurate representation of $x_i(t)$. Theoretically, if we have an infinite precision representation of $x_i(t)$, then the mean estimated error $\sqrt{E[(x-\hat{x})^2]}$ using the classical method would decrease by a factor of $\frac{1}{\sqrt{N}}$. However, an accurate representation of $x_i(t)$ implies that each sensor must be able to resolve a small difference in the measured data. For example, when measuring the temperature, a sensor node must be able translate temperatures to electrical signals in a fairly precise manner, e.g., 68.2F and 68.5F would result in two electrical signals of 0.90 volts and 0.92 volts, respectively. In addition, an accurate representation $x_i(t)$ requires a larger number of bits, whence more bits must be transmitted per measurement per sensor.

Our proposed framework eliminates the need for high resolution sensors. In particular, each sensor only makes a decision whether its measured data is greater or smaller than a certain threshold. This simplification allows one to build a sensor network consisting of cheap and simple sensors. This simplification also implies that each measurement is represented by only 1 bit, resulting in a reduction of transmission energy per node. Formally, we make the following assumptions:

1. Signal $x(t)$ is a random process with zero mean. The zero-mean assumption is for ease of analysis. It is not critical as any random process can be converted to a zero-mean process by simply subtracting its mean.

2. Additive noise $n_i(t)$ at each sensor is independent and identically distributed with zero mean. In other words, the additive noise samples at each sensor is independent and have the identical statistics.

3. Each sensor node is only able to detect the sign of $x_i(t)$, i.e., binary sensor.

We also omit the discussion of scheduling and routing protocols. These issues can be found in (14)(26).

Accuracy Consideration. *Our goal is to estimate* $x(t)$ *as* $\hat{x}(t) = f(y_1(t), ..., y_N(t))$ *for some function* $f(.)$*, where* $y_i(t) = sgn(x_i(t))$*, given the statistics of* $x(t)$ *and of the noise* $n_i(t)$. We do not make any assumption on temporal correlation of the data, i.e., $x(t-a)$ cannot be used to estimate $x(t)$. Thus, we shall omit the index t in all the variables, e.g., $x_i(t)$ will become x_i.

There are two important performance indicators for our sensor network: the Mean Square Error (MSE) and the Conditional Mean Square Error (CMSE) which are defined as

$$MSE \triangleq E[(x - f(y_1, y_2, ..., y_N))^2].$$
$$CMSE(m) \triangleq E[(x - f(y_1, y_2, ..., y_m))^2 | y_1, y_2, ..., y_m],$$

where $m \leq N$. The MSE characterizes the measured data accuracy of a sensor network, while $CMSE(m)$ characterizes the average amount of errors given a set of observations from m sensors. If fewer sensors are used, we would expect a larger $CMSE(m)$ value. MSE enables us to characterize the average performance of our sensor network in a certain environment. On the other hand, the $CMSE(m)$ is useful for adaptive data gathering to reduce the energy consumption. Also, we shall drop the index m from $CMSE(m)$ when the outputs from all the sensors are used.

Adaptive Data Gathering. To illustrate the energy reduction based on adaptive processing, we consider the following example. Suppose that a sensor network consisting of N nodes is designed to continuously measure a signal x and alarms the population whenever it detects $x > a$. In a real world scenario, $x > a$ may represent an anomaly such as the presence of an intruder or a dangerous chemical agent. Using all N data points, the processing node can estimate x as $f(y_1, y_2, ..., y_N)$ and determine whether $f(y_1, y_2, ..., y_N) > a$. Since $f(y_1, y_2, ..., y_N)$ is an estimate of x, it is also useful to determine the confidence level of this estimate in order to reduce the number of false alarms or to avoid missing an anomaly. As will be shown shortly, the accuracy of the estimate depends on the current realized value of x and on the number of data samples. Thus, if after a number of transmissions $m < N$, the intermediate nodes can estimate that $x < a$, i.e., no anomaly, with high confidence, then subsequent transmissions will be not be necessary, resulting in an overall energy reduction of the network.

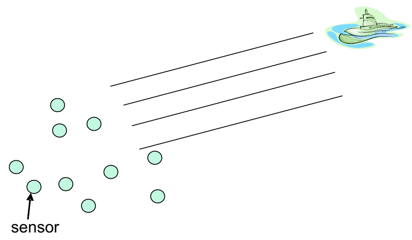


Figure 2: *Detection of an intruded submarine. Magnetic signatures at different sensors are almost identical due to the long distance of the submarine.*

Energy Efficient Coding. Energy consumption can be further reduced by employing our proposed coding scheme which exploits the network topology. The main idea is that, to estimate the signal, the processing node only needs to know the summary information such as the number of 1's. Thus, each node can reduce the number of transmitted bits accordingly based on its positions. We discuss this technique in detail in Section 5. We now provide the motivation for our proposed sensor network model.

3.2 Model Rationale

While our model can be used in different applications, it is primarily motivated by the characteristics of an intrusion detection application using a network of cheap magnetic sensors [1]. Currently, the coarse magnetic sensors can be made inexpensive to detect the presence of metallic materials at reasonably long distance, e.g., up 400 meters. Figure 2 shows a scenario in which a submarine is detected by a network of magnetic sensors at a long distance. Because of the long distance, the magnetic signatures (strength) at the sensors have approximate amplitude. Thus, $x(t)$ in our model above is assumed to be identical or highly correlated. An inexpensive binary magnetic sensor detects the presence of metallic materials by examining the polarity of the magnetic field, e.g., either up or down. However, in the typical situations, the thermal noise can flip its polarity, thus the measurements can be inaccurate. The thermal noise is an identically independent distributed (i.i.d)

[1]We are currently developing applications for these magnetic sensors at OSU.

noise and its affect on the magnetic strength is large as compared to the true magnetic signatures of the metallic objects. Therefore, we assume an i.i.d model for the noise $n(t)$. Furthermore, we will show that our sensor network will achieve good accuracy when the additive noise $n(t)$ is relatively large.

One approach to eliminate the network aspect for binary sensors is to put all the sensors on a single chip to make an accurate sensor. However, this approach lacks robustness since a single failure in its circuitry would potentially halt the function of the sensor. We note that these sensors are designed to operate under harsh conditions, and thus one must take into account the possibility of sensor failure. Another consideration is the security issue. In many applications, it is desirable to have the sensors deployed in geographically dispersed locations so as to make it harder for an attacker to disable the sensors. We now begin with the accuracy analysis of the proposed stochastic binary sensor network.

4 Accuracy Analysis

In this section, we analyze the accuracy of the MSE and $CMSE$ for the proposed binary stochastic sensor network operating in a typical environment where the additive noise is either uniformly or normally distributed. Since the MSE and $CMSE$ also depend on the distribution of the signal, we also characterize MSE and $CMSE$ for the signals having uniform distribution and normal distribution. We first begin with the analysis of our proposed sensor network in a uniform noise environment.

4.1 Uniform Noise

We assume that the noise samples at different sensors are independent and identically distributed. Let us consider the problem of estimating a signal $x \in [-1, 1]$ with additive uniform noise n_i over the interval $[-0.1, 0.1]$. If only one bit is used to determine the value of x at each sensor, then one possible scheme is for each sensor to output 1 if its measured data $x_i = x + n_i > 0$ and -1 otherwise. The processing node can then estimate x as the average of 1's and -1's from all the sensors. However, in this particular scenario, regardless of the number of sensors used in the estimation, a constant signal $x = 0.5$ will always be quantized to 1 because $x_i = x + n_i > 0$ for all i. Subsequently, the estimated output will always be equal to 1, which is 0.5 off the actual value. This problem results directly from having a small noise power. It is interesting to note that traditionally small noise, or, equivalently, high signal to noise ratio (SNR), leads to a better estimation of the signal. However, using a binary sensor, small noise carries no useful information due to the coarse quantization, thus resulting in a large MSE. In general, to achieve high accuracy using binary sensors in a uniform noise environment, the noise range must be equal to or larger than the range of the signals.

Given a signal x and the observations at different nodes $y_1, y_2, ..., y_N$, we must determine a good estimator $\hat{x} = f(y_1, y_2, ..., y_N)$ for x, where $y_i = sgn(x + n_i)$. While the optimal estimator for a random variable is the minimum mean square estimator (MMSE). The MMSE estimator, however, is complex since one needs to compute $E[X|Y = y]$ which requires an integration of the explicit conditional probability function over all possible value of x for every observed y[2]. Instead, when the signal and noise distributions are uniformly distributed over $[-\alpha, \alpha]$ and $[-\beta, \beta]$, respectively, and $\beta \geq \alpha$, we employ a linear estimator since this estimator uses only equation for all possible observed values which is suitable technique for simple sensor network. We have the following results.

Theorem 4.1. *The least square linear estimator $\hat{x}$ is*

$$\hat{x} = \frac{\gamma}{N}\sum_{i=1}^{N} y_i = \frac{\alpha^2\beta}{\alpha^2(N-1)+3\beta^2}\sum_{i=1}^{N} y_i. \tag{3}$$

and the corresponding MSE and $CMSE$ are:

$$MSE \triangleq E[(x-\hat{x})^2] = \frac{(3\beta^2-\alpha^2)\alpha^2}{3[\alpha^2(N-1)+3\beta^2]} \tag{4}$$

$$CMSE \triangleq E[(x-\hat{x})^2|\hat{x}] = \frac{A}{B}, \tag{5}$$

where

$$\begin{aligned} A = & \sum_{k=0}^{j}\sum_{i=0}^{N-j}\frac{(-1)^i}{\gamma^{i+k}}\binom{j}{k}\binom{N-j}{i}\alpha^{i+k+1} \\ & \times\left(\alpha^2\frac{1+(-1)^{i+k}}{i+k+3}+2\gamma\alpha\left(\frac{2j}{N}-1\right)\frac{(-1)^{i+k}-1}{i+k+2}\right. \\ & \left.+\gamma^2\left(\frac{2j}{N}-1\right)^2\frac{1+(-1)^{i+k}}{i+k+1}\right) \end{aligned} \tag{6}$$

and

$$B = \sum_{k=0}^{j}\sum_{i=0}^{N-j}\frac{(-1)^i}{\gamma^{i+k}}\binom{j}{k}\binom{N-j}{i}\frac{1+(-1)^{i+k}}{i+k+1}\alpha^{i+k+1}. \tag{7}$$

with j denoting the number of 1's in the outputs, and is directly related to $\hat{x}$.

Proof. Let $f_n(x)$ and $f(x)$ be the probability density functions of the noise and of the signal, respectively. Also denote the number of non-negative samples as j, then $\hat{x} = \gamma(\frac{2j}{N} - 1)$. We want to show that

$$\gamma = \frac{N\alpha^2\beta}{\alpha^2(N-1)+3\beta^2} \tag{8}$$

Now,

$$\begin{aligned} E[(x-\hat{x})^2|x] &= E[x^2|x] - 2E[x\hat{x}|x] + E[\hat{x}^2|x] \\ &= x^2 - 2x\gamma E\left[\left(\frac{2j}{N}-1\right)\middle|x\right] \\ &\quad +\gamma^2 E\left[\left(\frac{2j}{N}-1\right)^2\middle|x\right] \end{aligned} \tag{9}$$

[2]We can compute the MMSE estimate using numertical ntegration or summations of many variables. However, these solutions are still computational expensive

Given x, j is a binomial random variable, hence

$$E[j|x] = Nq, \tag{10}$$
$$E[j^2|x] = (Nq)^2 + Nq(1-q), \tag{11}$$

where q is the probability that $sgn(x+n_i) > 0$ which equals

$$q = \frac{1}{2}\left(\frac{x}{\beta} + 1\right). \tag{12}$$

Now, substitute $E[j|x]$ and $E[j^2|x]$ into Equation (9), we obtain

$$\begin{aligned} E[(x-\hat{x})^2|x] &= 4\gamma^2(Nq(1-q) + (Nq)^2)/(N^2) \\ &\quad -4\gamma^2 q + \gamma^2 - 4\gamma x q + 2\gamma x + x^2 \end{aligned} \tag{13}$$

Next, the mean square error taken over the input range $[-\alpha, \alpha]$ with the the uniform distribution density function $f(x) = 1/2\alpha$ is

$$\begin{aligned} MSE &= E[(x-\hat{x})^2] = \int_{-\infty}^{\infty} E[(x-\hat{x})^2|x] f(x) dx \\ &= \int_{-\alpha}^{\alpha} E[(x-\hat{x})^2|x] \frac{1}{2\alpha} dx \\ &= -\frac{\alpha^2\gamma^2}{3N\beta^2} + \frac{\alpha^2\gamma^2}{3\beta^2} - \frac{2\alpha^2\gamma}{3\beta} + \frac{\alpha^2}{3} + \frac{\gamma^2}{N} \end{aligned} \tag{14}$$

Now, to obtain the least square estimate, we take the derivative of (14) with respect to γ, and set it to zero to solve for γ which is equal to:

$$\gamma = \frac{N\alpha^2\beta}{\alpha^2(N-1) + 3\beta^2} \tag{15}$$

The corresponding MSE can be obtained by substituting (15) into (14).

The proof for $CMSE$ is provided in the Appendix. □

From Theorem 4.1, for large N, $\gamma \approx \beta$, thus the least square linear estimator depends only on the noise range, and not on the range of the signal. Theorem 4.1 also states that the MSE is inversely proportional to N. This result agrees with our intuition that a larger N leads to a smaller estimation error. Also, when a few sensors fail, the MSE will increase only slightly. In general, this approach enables a graceful degradation of data accuracy in the presence of failed sensors.

Figure 3(a) shows the MSE as a function of the number of sensors for three different uniform noises with different ranges: [-1, 1], [-1.1, 1.1], and [1.2, -1.2]. As seen, the MSE for the noise having the range $[-1, 1]$ is smallest. In fact, it is easy to show that the the minimum MSE is achievable when $\alpha = \beta$, i.e., the noise range equals precisely the signal range. Furthermore, the MSE becomes larger as the gap between the noise range and the signal range increases. One important observation is that the MSE asymptotically vanishes as the number of sensors N increases. This is important since we can guarantee an arbitrarily small estimation error, even in an arbitrarily large noise environment when using an appropriate number of sensors. Figure 3(a) shows this asymptotical decrease of MSE as the number of sensors increases.

Figure 3(b) shows the $CMSE$ as a function of the estimates. Clearly, the $CMSE$ depends on $\hat{x}$ ($\hat{x} = \gamma(\frac{2j}{N} - 1)$) and N. In particular, for a uniform signal and a uniform noise, the $CMSE$ increases when the magnitude of $\hat{x}$ decreases, and vice versa. Thus, to reduce energy consumption, a sensor can operate as follows. An intermediate sensor can compute $\hat{x}$ after collecting a number of samples from other sensors. Based on the current value of $\hat{x}$ and the corresponding estimation error $CMSE$, it can decide whether or not to continue relaying the data to next sensor. If the corresponding $CMSE$ is too high (low confidence level), the network would continue to collect more samples. On the other hand, if the $CMSE$ is small (high confidence level), the network will stop collecting data and thus reduce the amount of energy consumption. This decision is application dependent, and will be discussed in Section 6.

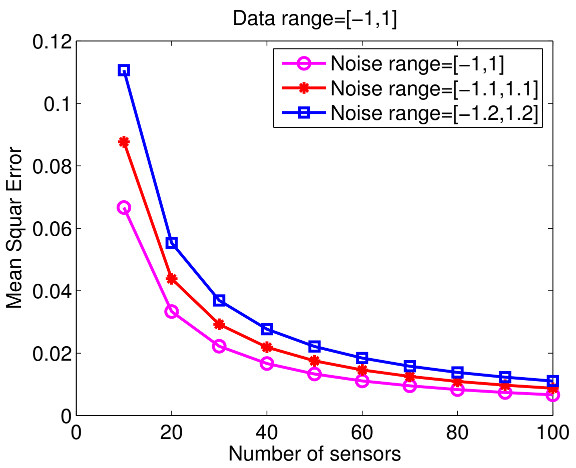
We note that computing the $CMSE$ is rather complicated; however, it can be computed once for all the possible values of $\hat{x}$ and the results are stored in a table at each sensor. When the adaptive data collection technique is used, each sensor can determine the corresponding $CMSE$ given $\hat{x}$ using a look-up table. We now consider a Gaussian signal in the presence of uniform noise. Similar to the uniform signal and uniform noise case, we assume that the range of the signal x is smaller than that of noise. However, this is not possible since the range of a Gaussian signal extends to infinity. Therefore, we assume that the signal standard deviation, σ, is much smaller than the range of the noise range, β in order for our binary sensor network to achieve reasonable performance. The following theorem characterizes the accuracy of our binary sensor network.

Theorem 4.2. *If the signal is normally distributed with mean 0 and variance σ^2, and the additive noise is uniformly distributed in the interval $[-\beta, \beta]$, with $\sigma << \beta$ and large N, then we have*

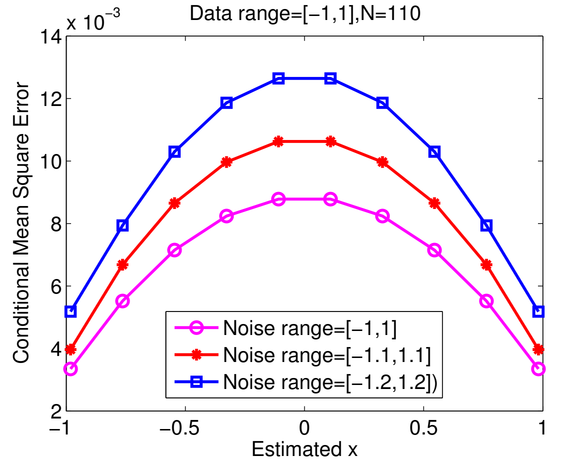
$$MSE \approx \frac{\beta^2 - \sigma^2}{N}, \tag{16}$$
$$CMSE \approx \frac{A}{B}, \tag{17}$$

where

$$\begin{aligned} A &= \sum_{k=0}^{j} \sum_{i=0}^{N-j} \frac{(-1)^i}{\beta^{i+k}} \binom{j}{k} \binom{N-j}{i} \sigma^{i+k+1} \frac{\sqrt{2\pi}}{2} \\ &\times \Bigg((i+k+1)!!\,\sigma^2 \left(1 + (-1)^{i+k}\right) \\ &- 2\beta \left(\frac{2j}{N} - 1\right) (i+k)!!\sigma \left(1 - (-1)^{i+k}\right) \\ &+ \beta^2 \left(\frac{2j}{N} - 1\right)^2 (i+k-1)!! \left(1 + (-1)^{i+k}\right) \Bigg) \end{aligned} \tag{18}$$



(a)



(b)

Figure 3: *Performance of a binary sensor network with uniform signal over* $[-1, 1]$ *under the different uniform noises. (a) MSE as a function of number of sensors; (b) CMSE as a function of* $\hat{x}$*, with the number of nodes* $N = 110$.

and

$$B = \sum_{k=0}^{j} \sum_{i=0}^{N-j} \frac{(-1)^i}{\beta^{i+k}} \binom{j}{k} \binom{N-j}{i} \sigma^{i+k+1} \times \frac{\sqrt{2\pi}}{2} (i+k-1)!!(1+(-1)^{i+k}) \quad (19)$$

with j *denoting the number of 1's in the outputs.*

Proof. When N is large, the least square linear estimator can be approximated as

$$\hat{x} = \beta \sum_{i=1}^{N} y_i. \quad (20)$$

Using (20) and after a few algebraic manipulations, we obtain

$$E[(x-\hat{x})^2|x] = \frac{\beta^2 - x^2}{N}.$$

Therefore,

$$E[(x-\hat{x})^2] = \int_{-\beta}^{\beta} \frac{\beta^2 - x^2}{N} f(x)dx, \quad (21)$$

where $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$. Since $\sigma << \beta$, we can approximate $\int_{-\beta}^{\beta} f(x)dx \approx \int_{-\infty}^{\infty} f(x)dx = 1$ and $\int_{-\beta}^{\beta} x^2 f(x)dx \approx \int_{-\infty}^{\infty} x^2 f(x)dx = \sigma^2$. Substituting these approximations into Equation (21), we obtain the desired MSE.

The proof regarding $CMSE$ is provided in the Appendix. □

Figure 4(a) shows the MSE as a function of the number of sensors for signal with standard deviation. $\alpha = 1$ under different noise ranges: [-4, 4], [-6, 6], -8, 8]. Similar to the uniform signal-uniform noise case, the MSE of the Gaussian signal and uniform noise vanishes as the number of sensors increases. Figure 4(b) shows the $CMSE$ as a function of the current estimates of the signal. Unlike in the uniform signal-uniform noise case, the $CMSE$ is small when $\hat{x}$ is small, and vice versa. As seen, a different assumption on signal distribution can lead to a very different curve for $CMSE$, which in turn affect the operations of the network. Therefore, the estimate must be designed carefully, taking into consideration of the signal and noise distributions.

4.2 Gaussian Noise

In many real-world scenarios, the noise often follows a normal distribution. Thus, it is important to characterize the MSE and $CMSE$ in these environments. While a linear estimator in Equation (3) results in high accuracy in an environment having an additive uniform noise, it is not optimal in the presence of additive Gaussian noise. We now derive an estimator for the signal x in the environment where the noise is normally distributed. First, let us consider the average of the quantized samples at the sensors

$$w_N = \frac{1}{N} \sum_{i=1}^{N} sgn(x + n_i). \quad (22)$$

When $N \to \infty$,

$$\begin{aligned} w \triangleq \lim_{N\to\infty} w_N &= \int_{-\infty}^{\infty} sgn(x+z) f_n(z)dz \\ &= -\int_{-\infty}^{-x} f_n(z)dz + \int_{-x}^{\infty} f_n(z)dz \\ &= 2\int_0^x f_n(z)dz, \end{aligned} \quad (23)$$

where $f_n(z)$ is the probability density function of the noise. For Gaussian noise, $f_n(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}}$, and therefore

$$w = erf(\frac{x}{\sqrt{2}\sigma}), \quad (24)$$

where

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

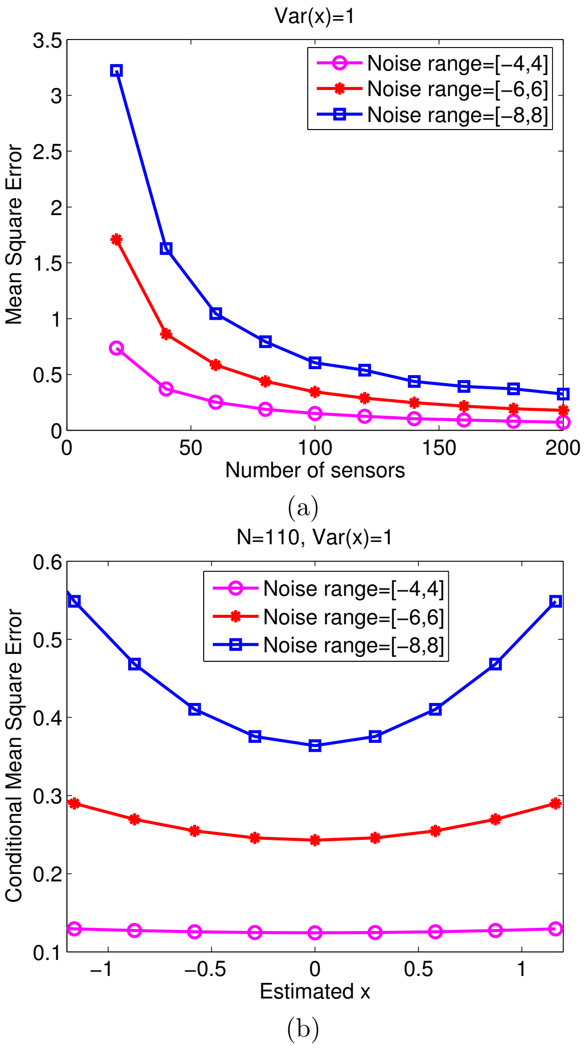


Figure 4: *Performance of a binary sensor network with signal $X \sim N(0,1)$ under different uniform noises. (a) MSE as a function of the number of sensors; (b) $CMSE$ as a function of $\hat{x}$, with number of nodes $N = 110$.*

Now, for a finite number of sensors N, a good estimate of x in terms of w is

$$\hat{x} = \sqrt{2}\sigma erf^{-1}(w) = \sqrt{2}\sigma erf^{-1}\left(\frac{2j}{N} - 1\right), \qquad (25)$$

where j is the number of non-negative samples. If the noise distribution has mean μ, then the estimator is simply:

$$\hat{x} = \sqrt{2}\sigma erf^{-1}\left(\frac{2j}{N} - 1\right) - \mu. \qquad (26)$$

Using the estimator in Equation (25), we obtain the following results:

Theorem 4.3. *If the signal is normally distributed with mean μ_s and variance σ_s and the additive noise is also normally distributed with mean μ and variance σ, then using the estimator in Equation (26), the MSE can be approximated as*

$$\frac{\sigma^2\sqrt{\pi}}{2N\sigma_s\sqrt{2}}\int_{-\infty}^{\infty}\left(1 - erf^2\left(\frac{x+\mu}{\sigma\sqrt{2}}\right)\right)e^{\left(\frac{(x+\mu)^2}{\sigma^2} - \frac{(x-\mu_s)^2}{2\sigma_s{}^2}\right)}dx \qquad (27)$$

and the $CMSE$ as

$$\frac{\int_{-\infty}^{\infty}(x-\hat{x})^2 q^j(1-q)^{N-j}e^{-\frac{(x-\mu_s)^2}{2\sigma_s{}^2}}dx}{\int_{-\infty}^{\infty} q^i(1-q)^{N-j}e^{-\frac{(x-\mu_s)^2}{2\sigma_s{}^2}}dx}, \qquad (28)$$

where

$$\hat{x} = \sigma\sqrt{2}erf^{-1}\left(\frac{2j}{N} - 1\right),$$

and

$$q = \frac{1}{2}erf\left(\frac{x+\mu}{\sqrt{2}\sigma}\right) + \frac{1}{2}.$$

Proof. The proof is provided in the Appendix. □

Figure 5(a) shows the MSE of a signal x having Gaussian distribution with stdev. $\sigma = 1$ as a function of the number of sensors for different Gaussian noises. As seen, the MSE vanishes as the number of sensors increases. Through simulations, we also observe that when the signal and noise have equal variances, the MSE appears to be minimum. Since we can prove this for the uniform signal and uniform noise case, we conjecture that the MSE is minimum whenever the signal and noise have identical distributions and parameters, regardless of the distribution types.

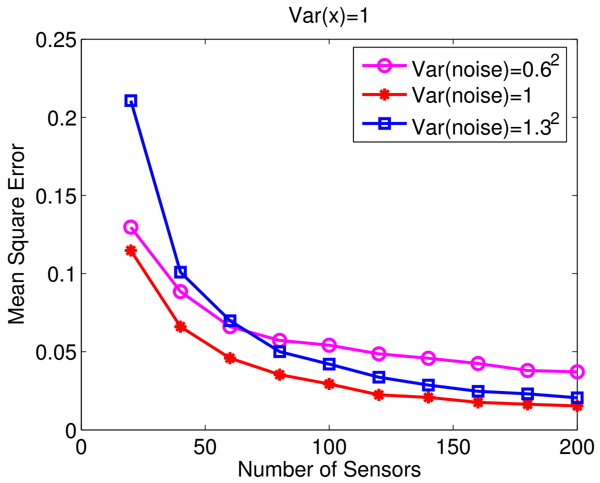
Figure 5(b) shows the $CMSE$ as a function of $\hat{x}$ under different Gaussian noises. Unlike the uniform signal-uniform noise case, the $CMSE$ is small when $abs(\hat{x})$ is small. This is intuitively plausible since the probability of having small x is large due to the signal having a normal distribution, and hence a small $\hat{x}$ is probably a better estimate of x. Again, knowing the signal and noise distribution is critical in designing the good estimate.

Finally, Figure 6 shows the $CMSE$ as a function of the number of sensor N for the signal and noise having identical normal distribution with $\sigma = 1$ and $\mu = 0$. Clearly, the larger N results in a smaller $CMSE$. In Section 6, we discuss in detail how an intermediate node can use the estimated $CMSE$ based on the current $\hat{x}$ and the number of sensors $m < N$ to save energy.

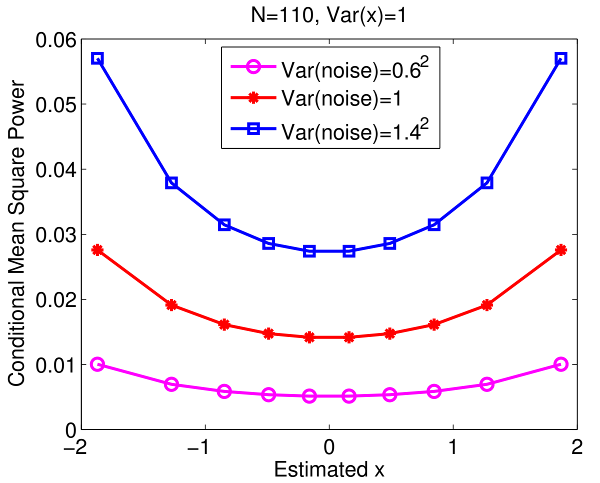
5 Energy Efficient Coding

In this section, we present the energy efficient coding. Most often, the energy saving is obtained through efficient routing. In this paper, we assume that the route for gathering data is already established. Our objective is to further improve energy efficiency through coding. We also assume that the energy consumption by the sensor nodes is proportional to the number of transmitted bits. Our objective is to minimize the number of bits sent in the network.

To illustrate our approach, we consider a sensor network consisting of four sensors arranged in a straight line as



(a)



(b)

Figure 5: *Performance of a binary sensor network with signal $X \sim N(0, 1)$ under the different Gaussian noises. (a) MSE as a function of the number of sensors; (b) CMSE as a function of $\hat{x}$, with number of nodes $N = 110$.*

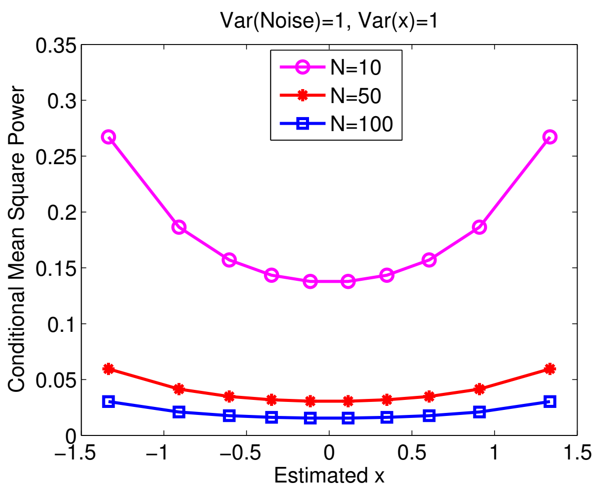


Figure 6: *CMSE as a function of $\hat{x}$ for different number of nodes N.*

shown in Figure 1. In our network, a sensor sends one bit per its measured sample $x(t)$. Depending on a sensor's position, it also relays many bits from other sensors. Thus, without coding, the total number of transmitted bits in this simple network is $1 + 2 + 3 = 6$ bits.

Now, one observes that the processing node 4 only needs to know the number of non-negative samples. In other words, to estimate x, node 4 does not need to know whether the measured values at each node is -1 or 1. Thus, instead of sending all 6 bits, node 3 may need to send to node 4 only 2 bits (4 possible patterns) to represent whether the number of non-negative samples is 0, 1, 2, or 3. Given this topology, it is impossible for node 4 to receive more than 4 non-negative samples. Similarly, node 2 needs to send only 2 bits to node 3 to indicate whether the number of non-negative samples is 0, 1, or 2. In general, by knowing its position in the topology, each sensor can code data adaptively to reduce the number of bits to be sent. We note that our technique is a variant of the *method of type* coding (27).

As an example, suppose that the quantized data measured at nodes 1, 2, 3 are -1, 1, and 1, respectively. Using our proposed coding scheme, sensor 1 would send a bit "0" to node 2 to indicate that no non-negative sample is observed. Sensor 2 then sends the bit pattern "01" to sensor 3 to indicate that 1 non-negative sample is observed so far. Node 3 then sends the bit pattern "10" to node 4 to indicate that the number of non-negative samples is now 2. Thus, the total number of bits sent in this case is $1 + 2 + 2 = 5$ bits, one bit fewer than non-coding approach. In general, assuming that the data is relayed according to the increasing order of node id, then a node n will need to send only $\lceil \log_2(n + 1) \rceil$ bits to node $n + 1$.

Although the previous example shows a modest energy reduction, for a sensor network consisting of a large number of nodes, this coding technique can result in substantial energy reduction. We have the following results:

Theorem 5.1. *Given N sensors arranged in a straight line with the processing node at one end, then the maximum number of bits which needs to be sent per sample in the stochastic binary sensor network with coding is:*

$$B_1 = N(m + 1) + 1 - 2^{m+1}, \tag{29}$$

and without coding is

$$B_2 = \frac{N(N - 1)}{2}, \tag{30}$$

where $m = \lfloor \log_2 N \rfloor$.

Proof. Assume that node N is the processing node. Without loss of generality, we have $2^m \le N < 2^{m+1}$, for $m = \lfloor \log_2 N \rfloor$. Then, the total number of bits sent by nodes 1 to $2^m - 1$ is

$$A_1 = \sum_{i=1}^{m} i 2^{i-1}. \tag{31}$$

Also, the total number of bits sent by the remaining nodes is

$$A_2 = (N - 2^m)(m+1). \quad (32)$$

Now, taking the derivative of $\sum_{i=1}^m x^i = \frac{x^{m+1}-1}{x-1} - 1$ with respect to x, we have

$$\sum_{i=1}^{m} i x^{i-1} = \frac{(m+1)x^m(x-1) - (x^{m+1}-1)}{(x-1)^2}. \quad (33)$$

Replacing $x = 2$, we have

$$A_1 = (m+1)2^m - 2^{m+1} + 1. \quad (34)$$

Adding A_1 and A_2, we obtain B_1. B_2 is easily obtained using arithmetic sum □

Theorem 5.1 indicates that when using coding in a straight-line topology, the energy consumption is on the order of $(N/\log N)$ times smaller than that required without coding. Note that one can use multiple straight-line topologies to construct a large network, e.g., a network of sensors on concentric circles with the data transmission taking place along the radius (straight-line) of these circles.

However, a straight-line topology may incur a high delay overhead. To reduce the delay, we propose to use a tree topology for data gathering. Figure 7 shows a binary tree topology for data gathering. Unlike the usual notation of a tree level, our tree level notation is reversed, namely, the leaf node is at level 1, the processing node is at level m. We prove the following theorem for data gathering in a tree with k branches. Data gathering starts from the nodes in the lowest to highest level. We have the following result for tree topology.

Theorem 5.2. *Using coding, the maximum number of transmitted bits per sample in a tree topology with k branches is*

$$B_k = k^{m-1} + \sum_{i=2}^{m-1} \left\lceil \log_2 \frac{k^i - 1}{k-1} \right\rceil k^{m+1-i}, \quad (35)$$

where m is the number of levels. For $k = 2$,

$$B_2 = 2(2^m - m - 1). \quad (36)$$

Proof. For an m-level tree, there are k^{m-1} leaf nodes. Each leaf node sends 1 bit of data. Therefore, the total number of bits sent by the leaf nodes is k^{m-1} bits. Now, each internal node needs to relay data for all its predecessors. If a node is at level $i > 1$, it has $\frac{k^i-1}{k-1} - 1$ predecessors (using the geometric sum). Since node i also needs to send 1 bit of its measured data, the maximum total number of coded bits sent by node i is $\left\lceil \log_2 \frac{k^i-1}{k-1} \right\rceil$. Finally, there are k^{m+1-i} nodes at level i, hence the total number of coded bits sent by all the nodes (excluding the processing node at level m) is

$$B_k = k^{m-1} + \sum_{i=2}^{m-1} \left\lceil \log_2 \frac{k^i - 1}{k-1} \right\rceil k^{m-i}. \quad (37)$$

For $k = 2$, we have

$$\begin{aligned} B_2 &= 2^{m-1} + \sum_{i=2}^{m-1} \left\lceil \log_2 (2^i - 1) \right\rceil 2^{m-i} && (38)\\ &= 2^m \sum_{i=1}^{m-1} i 2^{-i} && (39)\\ &= 2(2^m - m - 1). && (40) \end{aligned}$$

□

6 Simulation Results for Binary Sensor Networks

In this section, we characterize the trade-off between the energy consumption and the accuracy of the stochastic binary sensor networks for detecting anomalies through simulations. In particular, we consider two special topologies for simulations. The first topology is a straight-line topology consisting of 128 binary sensor nodes arranged in a straight line, with the processing node at one end. The measured data flows from one end of the line to the processing node at the other end. Data is accumulated along the way so that the processing node has all the measured data. The second topology is a tree topology consisting of 127 binary sensor nodes. Data is relayed from the leaf nodes to the internal nodes, and subsequently to the processing node as shown in Figure 7.

The main idea for reducing energy consumption in these networks is for a node to stop relaying data to the processing node if it determines with high confidence that the current estimated data is not anomalous. In particular, in these simulations, we consider a data point x anomalous if the estimated $abs(\hat{x}) \geq a$ and $CMSE = E[(x-\hat{x})^2|\hat{x}] < b$ where a and b are some threshold values set by the applications. Using this framework, each node would estimate the current data based on its own measurement and the measurements relayed to it from other nodes. We note that the $CMSE$ is employed in the decision making of a node to express the confidence level in the estimated data $\hat{x}$.

Using this model, a node in both straight-line and tree sensor networks operates as follows.

1. Initially, if a node is a leaf node in a tree topology or the first node in a line topology, it would send its data to the next node.

2. An internal node may send data only if it receives data from at least one node. This implies that, if all the predecessor nodes of a node determine that no further transmission is necessary, that node will honor the predecessor's decision.

3. If a node receives data from its predecessor node(s), it estimates the current value $\hat{x}$ and the $CMSE = E[(x-\hat{x})^2|\hat{x}]$ based on its own measurement and the relayed measurements from other node(s). If $abs(\hat{x}) < a$ and $CMSE < b$, it stops relaying data to the next node. Otherwise, it sends data to the next node.

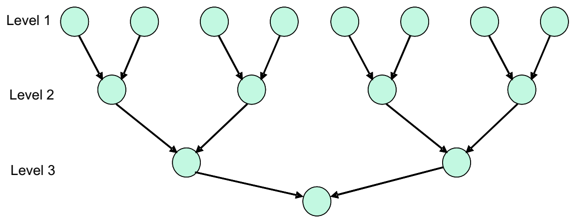


Figure 7: *Tree topology for data gathering.*

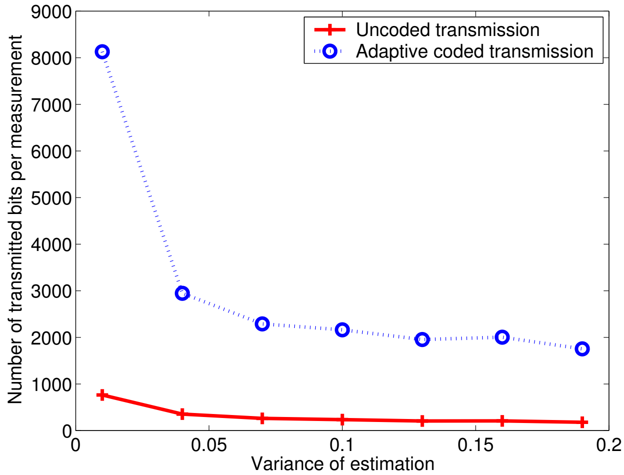
To characterize the energy reduction due to our proposed coding technique and the adaptive data collection, for each topology, we perform the simulations using coding and without coding. The measured signal x has a normal distribution with $\mu = 0$ and $\sigma = 1$. The additive noise is also normally distributed and has identical statistics to the signal. The threshold value a is set to 1, while b is varied to characterize the trade-off between the energy consumption and the data accuracy. Figure 8(a) shows the number of transmitted bits per data measurement as a function of $CMSE$ for the straight-line topology. It is noticed that, using coding reduces the number of transmitted bits approximately by a factor of 10 compared to without using coding. Also, if an application allows a larger estimation error, further energy reduction can be obtained, e.g., the number of transmitted bits with the $CMSE = 0.18$ is 8 times smaller than that of using the $CMSE = 0.01$.

Similarly, Figure 8(b) shows substantial saving of using coding in the tree topology. On the other hand, the adaptive data collection technique does not reduce the energy consumption as much. We note that the tree topology is much more energy efficient that the straight-line topology. This is because a bit in a tree topology does not have to be relayed many times as in the line topology.

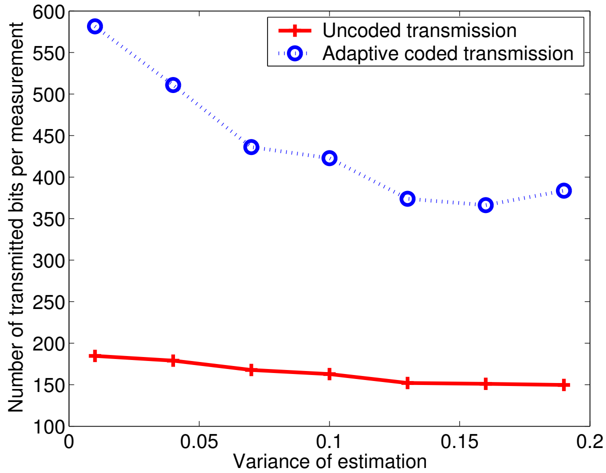
Although we present the simulation results for the line and tree topologies, we note that our coding technique can be applied to other topologies as well. The fundamental idea is to allow each sensor to code the data based on its position. When a sensor fails, the topology information must be disseminated to all the nodes to allow a new coding scheme. Even though sensor failures are allowed in our scheme, we believe that the failure frequency is small enough to warrant the overhead bandwidth for dissemination of topology information to all the sensors.

7 Conclusions

We have proposed a stochastic framework for detecting anomalies or gathering events of interest in a noisy environment using a sensor network consisting a large number of *cheap* binary sensors. We present the theoretical analysis of the accuracy of such sensor networks in different environments. We also propose an adaptive data collection framework based on the current measurements and a novel coding scheme in order to reduce the energy consumption. The simulation results of two stochastic binary



(a)



(b)

Figure 8: Number of transmitted bits per measurement as a function of $CMSE$ for (a) a straight line topology and (b) a tree topology.

sensor networks for anomaly detection using our proposed coding scheme and adaptive data gathering show that energy consumption can be reduced substantially, e.g., a factor of 10 for many scenarios.

REFERENCES

[1] I Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, 2002.

[2] J. Kahn, R. Katz, and K. Pister, "Mobile networking for smart dust," in *ACM/IEEE International Conference on Mobile Computing and Networking*, august 1999.

[3] J. Rabaey, J. Ammer, J. Silva, and D. Patel, "Picoradio:ad-hoc wireless networking of ubquituous low-energy sensor/monitor nodes," in *Proceedings*

of the IEEE Computer Society Annual Workshop on VLSI, april 2000.

[4] J. Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan, “Building efficient wireless sensor networks with low-level naming,” in *18th ACM Symposium on Operating Systems Principles*, october 2001.

[5] C. Intanagonwiwat and R. Govindan D. Estrin, “Directed difussion: A scalable and robust communication paradigm for sensor networks,” in *ACM/IEEE International Conference on Mobile Computing and Networks*, august 2000.

[6] W. Du, L. Fang, and P. Ning, “Lad: Localization anomaly detection for wireless sensor networks,” in *In The 19th International Parallel and Distributed Processing Symposium (IPDPS)*, April 2005.

[7] T. Vercauteren, D. Guo, and X. Wang, “Joint multiple target tracking and classification in collaborative sensor networks,” *IEEE Journal On Selected Areas in Communications*, vol. 23, pp. 714–723, 2005.

[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in *33rd International Conference on System Sciences*, january 2000.

[9] S. Servetto, “Distributed signal processing algorithms for the sensor broadcast problem,” in *Conference on Information Sciences and Systems*, march 2003.

[10] J. Barros and S. Servetto, “On the capacity of the reachback channel in wireless sensor networks,” in *Proceedings of IEEE International Workshop Multimedia Signal Processing*, 2002.

[11] A. Scaglione and S. Servetto, “On the interdependence of routing and data compression in multi-hop sensor networks,” *To appear in ACM/Kluwer Journal on Mobile Networks and Applications (MONET)*.

[12] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, “On network correlated data gathering,” in *Proceeding of INFOCOM*, 2004.

[13] S. Lindsey, C. Ragavendra, and K. Sivalingam, “Data gathering algorithms in sensor networks using energy metrics,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, pp. 776–787, 2002.

[14] C. Perkins, “Ad hoc on-demand distance vector (aodv) routing,” in *IETF MANET, Internet-Draft*, december 1997.

[15] J. Chang and L. Tassiulas, “Energy conserving routing in wireless ad-hoc networks,” in *Proceedings of INFOCOM*, 2000.

[16] D. Marco, E. Duarte-Melo, M. Liu, and year = 2003 D. Neuhoff title = On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data, booktitle = IPSN, ,” .

[17] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, “Network correlated data gathering with explicit communication: Np-completeness and algorithms,” *Submitted to IEEE/ACM Transactions on Networking*, 2004.

[18] P. Ishwar, A. Kumar, and K. Ramchandran, “On distributed sampling in dense sensor networks: a bit-conversation principle,” in *IPSN*, 2003.

[19] R. Cristescu and B. Beferull-Lozano, “Lossy network correlated data gathering with high-resolution coding,” *Submitted to IEEE/ACM Transactions on Network and IEEE Transaction on Information Theory, Special Issue on Networking and Information Theory*, March 2005.

[20] Z. Luo, “Lan isotropic univeresal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 735–744, April 2005.

[21] M. Enachescu, A. Goel, R. Govindan, and R. Motwani, “Scale free aggregation in sensor networks,” in *Proceedings of the First International Workshop on Algorithmic Aspects of Wireless Sensor Networks (AlgosensorsSOSP*, 2004.

[22] M. Sharaf, J. Beaver, A. Labrinidis, and P. Chrysanthis, “Balancing energy efficiency and quality of aggregate data in sensor networks,” *VLDB journal*, vol. 13, pp. 384–403, 2004.

[23] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based average consensus,” in *Proceedings of the 4th International Conference onInformation Processing in Sensor Networks*, April 2005, pp. 63–70.

[24] Z. Cvetkovic and I. Daubechies, “Single-bit oversampled a/d conversion with exponential accuracy in the bit-rate,” in *Data Compression Conference*, 2000.

[25] M. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, 1996.

[26] A. Woo and D. Culler, “A transmission control scheme for media access in sensor networks,” in *ACM/IEEE International Conference on Mobile Computing and Networks*, july 2001.

[27] T. Cover J. Thomas, *Elements of Information Theory, second edition*, Wiley, 2006.

Proof of Theorem 4.1

Proof. To obtain $CMSE$, let $f(x|j)$ be the conditional density function of x given j, the number of non-negative samples, and $f(x)$ be the signal's probability density distribution, then we have

$$\begin{aligned} E[(x-\hat{x})^2|\hat{x}] &= E[(x-\hat{x})^2|j] \quad (.41)\\ &= \int_{-\infty}^{\infty}(x-\hat{x})^2 f(x|j)dx \quad (.42)\\ &= \int_{-\infty}^{\infty}\frac{(x-\hat{x})^2 f(j|x)f(x)}{\int_{-\infty}^{\infty} f(j|x)f(x)dx}dx\\ &= \frac{\int_{-\alpha}^{\alpha}(x-\hat{x})^2 q^j(1-q)^{N-j}dx}{\int_{-\alpha}^{\alpha} q^j(1-q)^{N-j}dx}\end{aligned}$$

where q is the probability that $sgn(x+n_i)\geq 0$, which is equal to

$$q=\int_{-x}^{\infty}\frac{1}{2\gamma}dt=\frac{1}{2}\left(\frac{x}{\gamma}+1\right). \quad (.43)$$

Replacing q in Equation (.41), we have

$$E[(x-\hat{x})^2|j]=\frac{\int_{-\alpha}^{\alpha}(x-\hat{x})^2(1+\frac{x}{\gamma})^j(1-\frac{x}{\beta})^{N-j}dx}{\int_{-\alpha}^{\alpha}(1+\frac{x}{\gamma})^j(1-\frac{x}{\gamma})^{N-j}dx}. \quad (.44)$$

Now, letting $A=\int_{-\alpha}^{\alpha}(x-\hat{x})^2(1+\frac{x}{\gamma})^j(1-\frac{x}{\gamma})^{N-j}dx$ and $B=\int_{-\alpha}^{\alpha}(1+\frac{x}{\gamma})^j(1-\frac{x}{\gamma})^{N-j}dx$, then $E[(x-\hat{x})^2|j]=\frac{A}{B}$, Using binomial expansion, we have

$$\begin{aligned} A &= \sum_{k=0}^{j}\sum_{i=0}^{N-j}\frac{(-1)^i}{\gamma^{i+k}}\binom{j}{k}\binom{N-j}{i}\int_{-\alpha}^{\alpha}(x-\hat{x})^2x^{i+k}dx\\ &= \sum_{k=0}^{j}\sum_{i=0}^{N-j}\frac{(-1)^i}{\gamma^{i+k}}\binom{j}{k}\binom{N-j}{i}\alpha^{i+k+1}\\ &\quad\times\bigg(\alpha^2\frac{1+(-1)^{i+k}}{i+k+3}+2\hat{x}\alpha\frac{(-1)^{i+k}-1}{i+k+2}\\ &\quad+\hat{x}^2\frac{1+(-1)^{i+k}}{i+k+1}\bigg). \quad (.45)\end{aligned}$$

Replacing $\hat{x}=\gamma(\frac{2j}{N}-1)$, we obtain A in Equation (6). B in Equation (7) can be obtained in a similar manner. □

Proof of Theorem 4.2

Proof. For the $CMSE$, we note that the integrations are in the range of $[-\beta,\beta]$

$$E[(x-\hat{x})^2|j]=\frac{\int_{-\beta}^{\beta}(x-\hat{x})^2(1+\frac{x}{\beta})^j(1-\frac{x}{\beta})^{N-j}f(x)dx}{\int_{-\beta}^{\beta}(1+\frac{x}{\beta})^j(1-\frac{x}{\beta})^{N-j}f(x)dx}$$

Using binomial expansion on $(1+\frac{x}{\beta})^j(1-\frac{x}{\beta})^{N-j}$ and replacing $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$ in the expression above, we obtain the numerator:

$$A=\sum_{k=0}^{j}\sum_{i=0}^{N-j}\frac{(-1)^i}{\beta^{i+k}}\binom{j}{k}\binom{N-j}{i}\int_{-\beta}^{\beta}(x-\hat{x})^2x^{i+k}e^{-\frac{x^2}{2\sigma^2}}dx \quad (.46)$$

and the denominator:

$$B=\sum_{k=0}^{j}\sum_{i=0}^{N-j}\frac{(-1)^i}{\beta^{i+k}}\binom{j}{k}\binom{N-j}{i}\int_{-\beta}^{\beta}x^{i+k}e^{-\frac{x^2}{2\sigma^2}}dx. \quad (.47)$$

Given $\sigma << \beta$, we can approximate

$$\int_{-\beta}^{\beta}x^m e^{-\frac{x^2}{2\sigma^2}}dx\approx\int_{-\infty}^{\infty}x^m e^{-\frac{x^2}{2\sigma^2}}dx$$

$$=\frac{\sqrt{2\pi}}{2}(m-1)!!\sigma^{m+1}(1+(-1)^m).$$

Substituting the above approximation and $\hat{x}=\beta\left(\frac{2j}{N}-1\right)$ into Equations (.46) and (.47), we obtain the desired result. □

Proof of Theorem 4.3

Proof. To compute the MSE, from Equation (26), we have

$$\hat{x}=\sqrt{2}\sigma erf^{-1}\left(\frac{2j}{N}-1\right)-\mu. \quad (.48)$$

Denoting $\hat{x}=g^{-1}(w)$, we have

$$E[(x-\hat{x})^2|x]=E[(x-g(w))^2|x] \quad (.49)$$

We can approximate $g(w)$ by expanding at $w=g^{-1}(x)$

$$\begin{aligned}\hat{x} &\simeq g(g^{-1}(x))+(w-g^{-1}(x))\frac{dg(w)}{dw}\bigg|_{w=g^{-1}(x)}\\ &= x+\left(w-erf\left(\frac{x+\mu}{\sigma\sqrt{2}}\right)\right)\frac{\sigma\sqrt{2\pi}}{2}e^{\frac{(x+\mu)^2}{2\sigma^2}}. \quad (.50)\end{aligned}$$

Thus,

$$\begin{aligned} E[(x-\hat{x})^2|x] &= E\left[\left(w-erf\left(\frac{x+\mu}{\sigma\sqrt{2}}\right)\right)^2\frac{\pi\sigma^2e^{\frac{(x+\mu)^2}{\sigma^2}}}{2}\bigg|x\right]\\ &= \frac{\pi\sigma^2e^{\frac{(x+\mu)^2}{\sigma^2}}}{2}E\left[\left(w-erf\left(\frac{x+\mu}{\sigma\sqrt{2}}\right)\right)^2\bigg|x\right].\end{aligned}$$

Recall that $w=\frac{2j}{N}-1$, and

$$E[j|x]=Nq;$$
$$E[j^2|x]=(Nq)^2+Nq(1-q), \quad (.51)$$
$$(.52)$$

where q is the probability that the sign of a sample is 1,

$$q=\frac{1}{2}erf\left(\frac{x+\mu}{\sqrt{2}\sigma}\right)+\frac{1}{2}. \quad (.53)$$

After a few derivations, we have

$$E[(x-\hat{x})^2|x]=\frac{\sigma^2\pi}{2N}\left(1-erf^2\left(\frac{x+\mu}{\sigma\sqrt{2}}\right)\right)e^{\frac{(x+\mu)^2}{\sigma^2}}. \quad (.54)$$

Given the pdf of the signal: $f(x) = \frac{1}{\sigma_s\sqrt{2\pi}} e^{-\frac{(x-\mu_s)^2}{2{\sigma_s}^2}}$, the MSE will be:

$$\begin{aligned} E[(x-\hat{x})^2] &= \int_{-\infty}^{\infty} E[(x-\hat{x})^2|x] f(x) dx \\ &= \frac{\sigma^2\sqrt{\pi}}{2N\sigma_s\sqrt{2}} \int_{-\infty}^{\infty} \left(1 - erf^2\left(\frac{x+\mu}{\sigma\sqrt{2}}\right)\right) \\ &\quad \times e^{\left(\frac{(x+\mu)^2}{\sigma^2} - \frac{(x-\mu_s)^2}{2{\sigma_s}^2}\right)} dx. \end{aligned} \tag{.55}$$

Note that for special case $\sigma_s = 1$ and $\mu_s = \mu = 0$, the MSE becomes

$$MSE = \frac{\sigma^2\sqrt{\pi}}{2N\sqrt{2}} \int_{-\infty}^{\infty} \left(1 - erf^2\left(\frac{x}{\sigma\sqrt{2}}\right)\right) e^{\frac{(2-\sigma^2)x^2}{2\sigma^2}} dx. \tag{.56}$$

Finally, the $CMSE$ can be obtained using the same approach. □