Online Bayesian Data Fusion in Environment Monitoring Sensor Networks

Yang Dingcheng,1 Wang Zhenghai,2 Xiao Lin,1 and Zhang Tiankui3

1 Information Engineering School, Nanchang University, Nanchang 330031, China
2 Southwest Institute of Electronic Technology, Chengdu 610036, China
3 Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Yang Dingcheng; ydcxuanyuan@msn.com

Received 6 December 2013; Revised 19 February 2014; Accepted 4 March 2014; Published 3 April 2014

1. Introduction

Advances in wireless communication technologies together with the aggressive feature size scaling in VLSI circuits have enabled the massive deployment of ultrasmall, cost efficient, and low power sensor nodes, which have led to the blossom of wireless sensor networks (WSN) in a wide range of applications, such as environment monitoring, battlefield surveillance, health care, and home automation [1].

Common goal in most WSN applications is to reconstruct the underlying physical phenomenon (e.g., temperature and humidity), based on sensor observations [2, 3]. However, the context of WSN makes this task very challenging. First, each sensor is characterized by low power supply and limited computation and communication capabilities due to various design considerations such as small size battery, bandwidth, and cost. Second, the harsh environments where the sensor is deployed further exacerbate the reliability of sensors. Subsequently, ensuring the reliability of the data in WSN is going to be very challenging for WSN designers.

In WSN, sensor observations are exposed to various sources of errors during the course of sensing, processing, and communication. First of all, sensors may report readings with time-invariant bias known as systematic errors due to residual sensor calibration errors [4]. For example, in case of a light sensor, the biased readings can be incurred by sensor hardware or external factors such as dust particles on the protective lens of the sensor. After sensing and quantization, the data can still be disturbed by various soft errors on chips caused by thermal noise and cosmic ray radiations [5]. Afterward, the data samples again suffered from the wireless channel errors during the data reporting to sink node [6]. These kinds of errors are different in terms of severity, occurrence rates, and statistics.

Various error control strategies designed to control different errors for WSN have been investigated extensively in the literatures. Recently, schemes can calibrate single-sensor system biases by tuning each individual sensor based on ground truth information and advanced...
collaborative schemes can handle system biases for large scale of sensors which have been proposed in [7, 8], respectively. To tackle soft errors on chip and channel errors, majority of the existing methods have been turned to introduce redundancy to handle these kinds of errors individually. For example, error correction codes (ECC) and triple modular redundancy (TMR) [9] have been used widely to control soft errors on chips. Similarly, ECC together with automatic retransmission requests (ARQ) [10] have been applied to control wireless channel errors.

All the aforementioned algorithms address the reliable data reception problem by either introducing redundancies or fitting statistical models of the monitored phenomena but ignore the data transfer channel (DTC) error information. Due to the promise of considerable performance improvements as has been proved in decision fusion [11, 12], target tracking [13], and multiple accesses [14] scenarios, channel-aware related algorithms have attracted wide interest. In this paper, we focus on the reliable data reception from a sensor node that quantizes the monitored phenomenon into K levels, which is the basic problem for WSNs. We wish to design a decision device which chooses a quantization level for the current phenomenon such that the probability of a correct decision is maximized. Both the temporal correlation of a physical phenomenon and the error information of the DTC are modeled as prior input to the decision device.

The rest of the paper is organized as follows. Section 2 introduces the system model of data reception at the sink node from a sensor node. Section 3 models the error-prone DTC by a discrete K-ary input and K-ary output channel, where K is the number of quantization levels. Then, a discrete time series model is used to estimate the value of the monitored physical phenomenon and the estimation error is modeled as a normal distribution in Section 4. Section 5 gives some results that demonstrate the efficiency of the proposed algorithms. Finally, conclusions and future extensions are given in Section 6.

2. System Models and Problem Formulation

The problem to be solved including the network and channel-aware Bayesian model is first described formally.

2.1. Network Model. Figure 1 depicts the typical topology of WSN. Let \( M = \{m_1, \ldots, m_N\} \) be a finite set corresponding to N sensor nodes observing a common phenomenon \( J \). The gathered data sequences at the fusion node, from the observation sequences of sensor \( m_n \), are denoted by \( \{z_{nt}\}_{t=1}^{\infty} \), where \( t \) represents the time index and \( n \) is the node index. \( \{y_{nt}\}_{t=1}^{\infty} \) denotes the fusion decisions using the observations of all the sensor nodes in set \( M \).

**Definition 1.** The phenomenon \( J \) is a continuous stochastic process bounded at interval \([l, u]\) and \( J^t \) is the realization of \( J \) at \( t \)th transmission phase. The sensor digitalizes the interval \([l, u]\) uniformly into \( K \) quantization levels \( \theta = \{\theta_1, \ldots, \theta_K\} \), where \( K = 2^q \) (\( q = 1, 2, \ldots \)). Each quantization level represents a subinterval with size \( \Delta = (u - l)/K \). Using \( \theta^t_i \) denotes that \( \theta^t_i = \theta_i \) (\( i = 1, 2, \ldots, K \)), where \( \theta^t \) is the digitalized version of \( J^t \) at \( t \)th transmission phase.

![Figure 1: System model for distributed fusion.](image)

2.2. Bayesian Data Reliability Model. As shown in Figure 1 at time \( t \), the sensor node \( m_n \) samples and quantified the observation as phenomenon \( J^t \). After sensing and quantization, the observation is reported to the fusion node. After receiving the data reporting from sensor node \( m_n \), it is proposed to design a data detector that makes a decision on the received data and the error probability of the data transfer channel such that the probability of a correct decision is maximized. Using Bayes’ rule, the posterior probabilities can be expressed as

\[
P\left( \theta^t_i / z^n_t \right) = \frac{P\left( z^n_t / \theta^t_i \right) P\left( \theta^t_i \right)}{P\left( z^n_t \right)},
\]

where \( P(\theta^t_i / z^n_t) \) are the conditional probabilities of the phenomenon \( \theta^t_i = \theta_i \) at time \( t \), given the event that the gathered data at the fusion node from sensor node \( m_n \) at time \( t \) is \( z^n_t \). The decision criterion is based on selecting the subinterval corresponding to the maximum of the set of posterior probabilities. \( P(\theta^t_i / z^n_t) \), \( P(z^n_t / \theta^t_i) \) and \( P(\theta^t_i) \) are the conditional probabilities of the event that the gathered data at the fusion node from sensor node \( m_n \) at time \( t \) is \( z^n_t \), given phenomenon \( \theta^t_i = \theta_i \) at time \( t \), and \( P(z^n_t / \theta^t_i) \) are the priori probabilities which specify the data transfer channel. \( P(\theta^t_i) \) are the probabilities of the phenomenon \( \theta^t_i = \theta_i \) at time \( t \) and are also the priori probabilities that depended on the statistic property of the phenomenon. The denominator of (1) can be expressed as

\[
P\left( z^n_t \right) = \sum_{i=1}^{K} P\left( z^n_t / \theta^t_i \right) P\left( \theta^t_i \right).
\]

From (1) and (2), we observe that the computation of the posterior probabilities \( P(\theta^t_i / z^n_t) \) requires knowledge of the priori probabilities \( P(z^n_t / \theta^t_i) \) and \( P(\theta^t_i) \).
3. Data Transfer Channel Model for WSN

In Section 2, we have demonstrated that the computation of the posterior probabilities \( P(\theta_j^t/z_j^t) \) requires knowledge of the priori probabilities \( P(z_j^t/\theta_j^t) \), which are dependent on the statistic property of the data transfer channel between the sensor of the sensor node who samples the phenomenon \( J \) and the fusion node shown in Figure 2. In this section, we first investigate the error models for the cosmic ray radiation induced errors, thermal noise on chip induced errors, and wireless channel errors. Then, we propose a discrete \( K \)-ary input and \( K \)-ary output channel model to model the data transfer channel as shown in Figure 2.

3.1. Model for Radiation Induced Transient Error. When a radioactive particle strikes a semiconductor device, it causes a transient pulse that may alter the logical state of the struck node. The generated transient glitches on the circuit can be propagated through the circuit and reversed value gets latched into sequential circuit and becomes an error. Soft error rate (SER) of a circuit is exponentially related to the critical charge \( Q_{CRIT} \), which is the minimum charge required to cause a soft error and is proportional to system supply voltage:

\[
P_{se} \propto F \cdot A \cdot \exp \left( -\frac{Q_{CRIT}}{Q_S} \right).
\]

\( F \) is the neutron flux (i.e., radiation intensity), which is related to the environment where the device operates. \( A \) is the area of the circuit sensitive to particle strikes. \( Q_S \) is the charge collection efficiency of the device. According to [5]

\[
Q_{CRIT} \propto \frac{V_{dd}}{\gamma}.
\]

\( V_{dd} \) is system supply voltage. As shown in (4), when supply voltage decreases the critical charge decreases, which will increase the transient error rate. Besides, the flux of low energy particles is orders of magnitude higher than high energy particles [15]. Therefore, with smaller critical charge, circuits are more vulnerable to transient errors induced by lower energy particle strikes.

3.2. Model for Thermal Noise Induced Transient Error. Thermal noise in logic gates is a stationary Gaussian stochastic voltage fluctuation process with zero mean value and deviation \( \delta_{\sigma} = \kappa T \), where \( \kappa \) is the Boltzmann constant and \( T \) is the temperature [16]. The model of logic gate considered herein is shown in Figure 2 and was first used by author [17]. It consists of three cascaded stages. The first stage is the logic function, which is used to compute the true value of the logic gate. The next stage is the additive white Gaussian noise (AWGN) channel, which is the noise channel when the logic value transfers through the logic gate. The essence of this stage is to add thermal noise to the data transferred from the input to output. The third is a threshold function, which restores binary data readout from the logic gate. Whenever the noise voltage exceeds the threshold value \( V_{th} \) of the logic gate, an error will happen. Clearly, the logic gate, through which data is transferred, can be modeled as a binary symmetric channel (BSC) [5]. The bit error rate (BER) of the BSC channel is \( P_c \):

\[
P_{noise} = \frac{1}{2} \text{erfc} \left( \frac{V_{th}}{\sqrt{4kT}} \right),
\]

where \( \text{erfc}(\gamma) = (2/\sqrt{\pi}) \int_\gamma^\infty \exp(-t^2) \, dt \) and we assume \( V_{th} = (1/2)V_{dd} \), which is the typical case in [5].

3.3. Model for Wireless Channel Error. Since most of the envisioned applications for WSNs assume that sensor nodes are densely lying on the ground and their antennas are a few centimeters over the ground, this near ground feature makes the channel models used in mobile wireless communication networks, where antennas are always over the ground by 1.5 meters. Many field measurement campaigns have verified and reported that one and two slope log-distance path loss models are suitable for near ground channel for WSNs deployed in outdoor scenarios [18]. Comparing to one-slope log-distance path loss model, two-slope log-distance path loss model can provide lower model error at 868 MHZ in ISM band. Because most of the radio energy is intercepted by the ground within the Fresnel zone, the path loss exponent will be bigger than the path loss exponent beyond the Fresnel zone. However, considering the small size of the Fresnel zone, we think one-slope log-distance path loss, which can be fitted by the measurements beyond the Fresnel zone, can accurately model the channel for WSNs:

\[
L(d) = \text{PL}(d_0) + 10\eta \log_{10} \left( \frac{d}{d_0} \right) + X_{\sigma},
\]

where \( \text{PL}(d_0) \) is the path loss at a reference distance \( d_0 \) in dB, \( \eta \) is the path loss exponent, and \( X_{\sigma} \) denotes the shadowing fading component, with \( X_{\sigma} \sim N(0, \sigma) \).

Definition 2. Equation (6) can accurately model the one-hop channel between two sensors within WSNs. \( h(d) \)
is the one-hop channel response in frequency domain for the two sensors at distance $d$ apart.

Based on Definition 2, we define an intermediate variable called channel fading factor $h'(d) = 10 \log_{10}(|h(d)|)$, which measures the amplitude of $h(d)$ in dB:

$$h'(d) = L(d) + 10\eta \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma.$$  \hspace{1cm} (7)

**Definition 3.** The additive Gaussian noise at the receiver is $n \sim N(0, \sqrt{N_0}/2)$. We further assume that the sensor nodes transmit binary PAM signals, where the two signals waveforms are $s_1 = \sqrt{\epsilon_b}$ and $s_2 = -\sqrt{\epsilon_b}$ in the signal duration intervals and zero elsewhere. The two signals are equally likely to be transmitted.

**Theorem 4.** If Definitions 2 and 3 hold, the probability of error between one hop is

$$P_{\text{hop}} = Q\left(\sqrt{\frac{2\|h\|^2\epsilon_b}{N_0}}\right).$$  \hspace{1cm} (8)

**Proof.** Let us assume $s_1$ is transmitted. Then, the received signal from the demodulator is

$$r = \|h\|^2 s_1 + h^*n = \sqrt{\epsilon_b} + n',$$

where $\epsilon'_b = \|h\|^2\epsilon_b$, $n' \sim N(0, \sqrt{N'_0}/2)$, $N'_0 = ||h||^2 N_0$. The optimal decision is made by the following rules:

$$p\left(\frac{r}{s_1}\right) \geq p\left(\frac{r}{s_2}\right) \quad r = s_1$$

$$p\left(\frac{r}{s_1}\right) < p\left(\frac{r}{s_2}\right) \quad r = s_2.$$  \hspace{1cm} (10)

The two conditional pdfs of $r$ are

$$p\left(\frac{r}{s_1}\right) = \frac{1}{\sqrt{\pi N'_0}} \exp\left(-\frac{(r - \sqrt{\epsilon'_b})^2}{N'_0}\right).$$  \hspace{1cm} (11)

$$p\left(\frac{r}{s_2}\right) = \frac{1}{\sqrt{\pi N'_0}} \exp\left(-\frac{(r + \sqrt{\epsilon'_b})^2}{N'_0}\right).$$

Given that $s_1$ is transmitted, the probability of error is simply the probability that $r < 0$; that is,

$$p\left(\frac{r}{s_1}\right) = \int_{-\infty}^0 p\left(\frac{r}{s_1}\right) dr = \frac{1}{\sqrt{\pi N'_0}} \int_{-\infty}^0 \exp\left(-\frac{(r - \sqrt{\epsilon'_b})^2}{N'_0}\right) dr$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{\epsilon'_b}/N'_0} \exp\left(-\frac{x^2}{2}\right) dx = Q\left(\sqrt{\frac{2\epsilon'_b}{N'_0}}\right).$$  \hspace{1cm} (12)

Similarly, if we assume that $s_2$ is transmitted, the received signal from the demodulator is

$$r = \|h\|^2 s_2 + h^*n = -\sqrt{\epsilon_b} + n'.$$  \hspace{1cm} (13)

Given that $s_2$ is transmitted, the probability of error is simply the probability that $r > 0$; that is,

$$p\left(\frac{r}{s_2}\right) = \int_0^\infty p\left(\frac{r}{s_2}\right) dr = \frac{1}{\sqrt{\pi N'_0}} \int_0^\infty \exp\left(-\frac{(r + \sqrt{\epsilon'_b})^2}{N'_0}\right) dr$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\epsilon'_b}/N'_0}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = Q\left(\sqrt{\frac{2\epsilon'_b}{N'_0}}\right).$$  \hspace{1cm} (14)

**Theorem 5.** Figure 3 shows the model for data transmission over an end-to-end link in $F$-hop WSNs with $F \geq 2$. If Definitions 2 and 3 hold, the bit error probability of the DTC in $F$-hop WSNs is

$$P_{\text{hop}} \leq Q\left(\sqrt{\frac{2\|h\|^2\epsilon_b}{N_0\left(1 + \sum_{f=1}^{F-1} \|h_f^f/h_f^f\|^2\right)}}\right).$$  \hspace{1cm} (16)
where $h_f$, $s_f$, and $r_f$ denote the channel response of the $f$th hop, the transmitted, and received signal by the $f$th relay node in the $t$th transmission phase, respectively. Particularly, $s_f$ and $r_f$ are, respectively, the transmitted signal by the sensor node and received signal at the sink node in the $t$th transmission phase.

Proof. First, we define that the $f$th relay node scales the received signal by a factor $g_f = 1/\|h_f\|$ at the transmission phase $t$. As shown in Figure 3, the received signal at the sink node in the $t$th transmission phase is $r_f = h_f s_f + n_0$ and the transmitted signal of this hop in this transmission phase is $s_{f-1} = g_{f-1} r_{f-1}$; then,

$$r_f = h_f g_{f-1} r_{f-1} + n_0. \tag{17}$$

The received signal of the first relay node in the $t$th transmission phase is

$$r_1 = h_1 g_0 + n_0. \tag{18}$$

Based on (17) and (18), the received signal at the sink node in the $t$th transmission phase is

$$r_f = s_f \prod_{j=1}^m h_j g_{j-1} + n_0 \prod_{j=2}^m h_j g_{j-1} + n_0. \tag{19}$$

where $g_0 = 1$ plug the scale factor $g_f$ into (19); the received signal at the sink node in the $t$th transmission phase can be expressed as

$$r_f = s_f \prod_{j=1}^m h_j + n_0 \prod_{j=2}^m h_j + n_0. \tag{20}$$

Clearly, the second term of (20) can be upper-bounded by

$$n_0 \prod_{j=2}^m h_j \leq n_0 \prod_{j=2}^{m-1} h_j. \tag{21}$$

Based on (20) and (21), using LS demodulator, the bit error probability of $F$-hop WSNs is (16).

4. Data Source Model for Physical Phenomenon

From (1) and (2) in Section 2, we observe that the computation of the posterior probability $P(\theta_j | z^n_j)$ requires knowledge of the priori probabilities $P(\theta_j)$, which are dependent on the statistic property of the phenomenon $J$. In this section, we first investigate methods to model the data source of the phenomenon $J$. Then, methods to update the parameters of data source models online are provided to ensure robustness.

Consider an environment monitoring scenario where the sink node needs to continuously reconstruct the data field based on the data gathered from all the sensor nodes spatially and temporally correlated. That is data at any given point in such a field are correlated not only with the data at nearby points but also with the previous values of the data measured at the same point.
4.1. Empirical Model for Data Source. Models to explore the spatial and temporal correlation of a physical phenomenon are investigated in [13]. As in [13], the quantization is assumed to be very fine, and the quantization error is ignored. If at time $t_2$, the sink node uses the data value gathered at the point $x_1$ and time $t_1$ to estimate the current value at point $x_2$, the corresponding mean square error is assumed as follows:

$$D(x_1, t_1; x_2, t_2) = 2 - 2 \exp \left( -\alpha(x_1 - x_2)^2 - \beta(t_1 - t_2)^2 \right).$$ \hspace{1cm} (25)

In (25), the spatial distortion is related to $|x_1 - x_2|$, and the temporal distortion is related to $|t_1 - t_2|$. $\alpha$ and $\beta$ are the spatial and temporal correlation parameters, and higher values of these parameters specify weaker correlated field. In this section, we focus on the prediction of the current value $\theta^i$ using the historically gathered data $\{\theta^{i-r}\}_{r=1}^{\tau}$. In the same location and make the following definitions.

**Definition 6.** Define the prediction function as $F(\cdot)$. The estimation of the current value $\theta^i$ is denoted by $\hat{\theta}^i$, and then

$$\hat{\theta}^i = F \left( \{\theta^{i-r}\}_{r=1}^{\tau} \right).$$ \hspace{1cm} (26)

We further define the $\ell_i = \theta^i - \hat{\theta}^i$ as the prediction error.

Time series models $F$ are used for the prediction of the future behavior of variables. These models account for the fact that observations have an inherent structure (e.g., autocorrelation, trend, or seasonal variation) that should be accounted for. Two commonly used forms of these models are autoregressive models (AR) and moving average (MA) models.

**Definition 7.** The distribution of the prediction error $\ell_i$ is a normal distribution, with $\ell_i \sim N(0, \delta^i)$.

Based on Definitions 4 and 5, we can give the data source model for the phenomenon $J$ as shown in Figure 5. Denote $P_0(\theta_j^i/\theta_j^k)$ as the probability of occurrence of the event that the current data gathered from sensor node $m_n$ is $\theta_j^i$ under the condition that $\theta_j^i = \theta_j^k$.

$$P_0 \left( \frac{\theta_j^i}{\theta_j^k} \right) = \frac{1}{\sqrt{2\pi} \delta} \exp \left( \frac{- (x - \theta_j^k)}{2 \delta^2} \right) dx$$

$$= \frac{1}{\sqrt{2\pi} \delta} \exp \left( \frac{z^2}{2} \right) dz$$

$$= \Phi \left( \frac{\theta_j^i - \theta_j^k + \Delta/2}{\delta} \right) - \Phi \left( \frac{\theta_j^i - \theta_j^k - \Delta/2}{\delta} \right)$$ \hspace{1cm} (27)

Given $\theta_j^{i-1} = \theta_j^k$, we observe that the parameter $P_0(\theta_j^i)$ of the Bayesian model (1) is equal to $P_0(\theta_j^i/\theta_j^{i-1})$ as shown in Figure 5. This empirical model is characterized by one parameter $\delta$ to be determined. Comparing to (25), we know that $\delta \propto \beta(t_1 - t_2)^2$, where $|t_1 - t_2|$ denotes the sample interval. In the runtime stage, we can update the parameter $\delta$ according to the observations gathered at the fusion node as $\delta = \max(|z_n^i - z_{n-1}^i|)$, with $\tau = t - w, \ldots, t - 1$. $w$ is the window length and is less than $t$.

4.2. Learning Model for Data Source. Beside the empirical models, we can also learn the parameter $P(\theta_j^i)$ in the Bayesian model (1) by random experiments. That is, the fusion node needs to keep $K^2$ counters to log the occurrences of the pair $(\theta_j^{i-1}, \theta_j^i)$. Learning the model parameter by training is simple. However, enough random experiments are required until the learning is reasonable.

5. Experimental Evaluations

5.1. Evaluation Setup. The data are quantized as 8-bit values in our evaluations, including sensor data about water temperature, dissolved oxygen in river water, and river stage from the California Data Exchange Center (CDEC) [19]. As shown in Figures 6, 7, and 8, these data sets differ in terms of autocorrelation properties and degrees of stationarity.

Most of the envisioned applications of WSNs assume that sensor nodes are densely lying on the ground and their antennas are a few centimeters over the ground. This near ground feature makes the channel in WSNs to be very different from the channel in mobile communication networks, where antennas are always over the ground by 1.5 meters. Many field measurement campaigns have verified and reported that the two-slope log-distance path loss model can accurately model the near ground channel in WSNs deployed in outdoor scenarios at ISM frequency band [20].
Table 1: Dataset parameters and the improvement factors of the proposed channel-aware Bayesian mode.

| Data set | Location       | Sampling period | Sensor type                     | Sample number | Data model | $\epsilon_{\text{in}}/\epsilon_{\text{out}}$ |
|----------|----------------|-----------------|--------------------------------|---------------|------------|---------------------------------|
| 1        | Lewiston       | 1 day           | Water temperature              | 181           | RHW        | 1.5                             |
| 2        | Balls ferry bridge | 1 hour         | Dissolved oxygen in river      | 2191          | AR, 10-order | 2.0                             |
| 3        | Rumsey bridge  | 1 hour          | River stage                    | 2448          | AR, 4 orders | 2.1                             |

Considering the small Fresnel zone, we assume there are no sensor nodes locating within the Fresnel zone of the other sensor nodes, and we can use one-slope path loss model shown in (28) as the channel between the transmitter and receiver of every hop:

$$20 \log(h) = L_0 + 10n \log\left(\frac{d}{d_0}\right) + X_\sigma,$$

(28)

where $d_0$ is the breakpoint, $L_0$ is the path loss at $d_0$ m, $n$ is the path loss factor, $d$ is the distance between the transmitter and receiver, and $X_\sigma$ is a normal random variable with standard deviation $\sigma$.

In experiments, we assume $d_0 = 2$ m, $L_0 = 70$ dB, $n = 4$, $\sigma = 6$ dB, the output power of the transmitter is $-30$ dBm, the noise power at the receiver is $N_0 = -174$ dBm, and there are no gains for both the antennas of the transmitter and receiver. We further assume that the relay nodes and sink node know perfect channel state information. The residual errors in the data after error corrections at the sink node are evaluated in terms of the normalized mean square error (NMSE) as follows:

$$\epsilon_{\text{out}} = \sqrt{\frac{1}{N} \sum_{v=1}^{N} \left(\theta_v - \hat{\theta}_v\right)^2}.$$

(29)

In (29), $N$ is the total number of samples. An AR model is first fitted by an offline process with the Yule-Walker method [21] in our evaluations. If the order or prediction error of the fitted AR model is too large, a RHW model [22] will be used instead. At the runtime stage, the parameters of the AR and RHW models are updated by each window of observations with the Yule-Walker method [21] and the method detailed in [22], respectively. The residual errors $\epsilon_{\text{out}}$ are compared with the NMSE $\epsilon_{\text{in}}$ of the error-corrupted data received at the sink node. For example, we list the improvement factors ($\epsilon_{\text{in}}/\epsilon_{\text{out}}$) of the proposed channel-aware Bayesian model in Table 1, where the sensor node directly routes the observations of the monitored phenomenon $J$ to the sink node and the distance between the sensor node and sink node is 40 m. In Table 1, we also give the parameters of the data sets used in our evaluation.

5.2. Evaluation in One-Hop WSNs. In this subsection, we evaluate the error resilience of the proposed Bayesian model, Reed-Solomon (RS) coding, and discrete time series models in one-hop WSNs, where the sensor node directly routes the observations of the monitored phenomenon $J$ to the sink node. And the distance between the sensor node and the sink node varies from 20 m to 100 m. Figures 9, 10, and 11 compare the NMSE of the data estimated by the proposed Bayesian model and Reed-Solomon (RS) coding, and the data estimated by discrete time series models (i.e., RHW or AR models) with the NMSE of the error-corrupted data received by the sink node. In Figures 9, 10, and 11, NMSE floors of the data estimated by discrete time series models can soon be noticed, which are dominated by estimation lags around relatively fast changes of the physical phenomena. The error-corrupted data received at the sink node and data corrected by Bayesian model and RS coding show NMSE floors as well; however, these NMSE floors are much lower than the error floor of the data estimated by RHW or AR models and the common source of these NMSE floors is quantization errors. From Theorem 4 and (28), we know that the bit error probability of one-hop transmission is directly proportional to the hop distance. Due to error correction ability, the RS coding exceeds the proposed Bayesian model and discrete time series models at very low channel error levels. However, this trend is reversed in the presence of high channel error levels, because large NMSE incurred by
the untreated decoding errors of RS coding in this case. The other side effect of RS coding is high communication overhead. For example, RS coding with 0.43 code rate ($R = 0.43$), which can nearly correct all errors at very low channel error levels, will increase the communication overhead by 57%. Unlike FEC, the overhead of the proposed Bayesian model is computation increasing at the sink node, which is caused by probabilities multiplications and can be solved by log-domain additions.

Depending on the channel error levels, the channel-aware Bayesian model will measure the quantization level with higher likelihood by more confidence. Particularly, at very low channel error levels (e.g., the distance of the hop is 20 m), the quantization levels with small distances (i.e., 0 or 1) to the channel output will have significant chance to be believed as the channel input by the channel-aware Bayesian model. Hence, at very low channel error levels the channel-aware Bayesian model performs much better than the AR and RHW models that only explore temporal correlations among observations.

5.3. Evaluation in Multihop WSNs with Amplify-and-Forward Relay Nodes. In this subsection, we evaluate the error resilience of our model, RS coding, and discrete time series models in multihop WSNs, where the readings of the monitored phenomenon $\theta$ observed by the sensor node are routed to the sink node by $f$ ($f = 1, 2, 3, 4$) relay nodes. The distance between the sensor node and the sink node is 200 m, and the distance of every hop is $200/(f + 1)$ m. Figures 12, 13, and 14 compare the performance of our model with RS coding and discrete time series models. In the multihop sensor networks, where the former relay node amplifies its received signal and forwards it to the next relay node or sink node, the bit error probability of the DTC for this scenario is given by Theorem 5. Clearly as shown in both Theorem 5 and experiment results, the reliability of the end-to-end link in multihop networks degrades considerably compared with the one-hop networks with the same average distance per hop, because the amplified noise in every relay node along the transmission will finally be cumulated at the sink node. However, given a field for the coverage of WSNs, deploying more AF relay nodes still can improve the reliability of the data reception at the sink node. Similar to the results in one-hop networks, our model is still superior over the AR or RHW models and performs as well as RS coding in all cases. However, because of the increased noise level at the sink nodes in multihop networks with AF relay nodes, the performance gain of our model compared with RS coding at medium to large hop distance is increased and the
6. Conclusion and Future Works

In this paper, we have presented a channel-aware Bayesian model to reliably recover data from transient errors on the data transfer channel between the sensor node which monitors a phenomenon and the sink node, and a discrete K-ary input and K-ary output channel has been provided to formulate the error information on the data transfer channel. Using the real sensor data from CDEC, we have evaluated the performance of our method in both one-hop and multihop sensor networks. In all scenarios, the channel-aware Bayesian model is obviously superior over the discrete time series models (i.e., RHW or AR models), which merely explore the correlations among the observations of the monitored phenomena. Furthermore, the proposed channel-aware Bayesian model performs as well as the RS coding with code rate 0.43; however, the later will increase the communication overhead by 57%. Besides, RS coding can only control the wireless channel errors. The shortcoming of the channel-aware Bayesian model, which is the computation increasing caused by probability multiplications at the sink node, can be solved by log-domain algorithms.

We have identified several future research avenues. First, we currently assume that the sink node and relay nodes know perfect channel state information, and in-field model performance still needs future evaluation. Moreover, the combinations of the proposed algorithms with distributed data fusion and signal detection algorithms considering noise at the sensor of individual sensor node are the most important extensions in future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of nos. 61250005, 61340025, 54th China Postdoctoral Science Foundation funded project (no. 2013M541875), Jiangxi postdoctoral Merit-funded project funds 2013KY07, and JiangXi Natural Science Foundation 20132BAB211055, and Foundation of Jiangxi Educational Committee (GJJ13062).

References

[1] Z. H. Wang, M. Tian, and Y. H. Wang, “Channel-aware Bayesian model for reliable environmental monitoring sensor network,” Journal of Communications, vol. 6, no. 5, pp. 355–359, 2011.
[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” IEEE Communications Magazine, vol. 40, no. 8, pp. 102–114, 2002.
[3] Y. Zhang, N. Meratnia, and P. Havinga, “Outlier detection techniques for wireless sensor networks: a survey,” IEEE Communications Surveys and Tutorials, vol. 12, no. 2, pp. 159–170, 2010.
[4] J. F. Sanford and B. A. Hamilton, “On-line sensor calibration and error modeling using single actuator stimulus,” in Proceedings of the IEEE Aerospace Conference, pp. 1–11, 2009.
[5] S. M. Jahinuzzaman, M. Sharifkhani, and M. Sachdev, “An analytical model for soft error critical charge of nanometric SRAMs,” IEEE Transactions on Very Large Scale Integration Systems, vol. 17, no. 9, pp. 1187–1195, 2009.
[6] Z. Li, C. Tao, and X. D. Zhang, “Distributed estimation for sensor networks with channel estimation errors,” Tsinghua Science and Technology, vol. 16, no. 3, pp. 300–307, 2011.
[7] C. M. Vuran and I. F. Akyildiz, “Cross-layer analysis of error control in wireless sensor networks,” in Proceedings of the 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON ’06), pp. 585–594, 2006.
[8] B. Zarei, V. Muthukumarasay, and X. W. Wu, “A residual error control scheme in single-hop wireless sensor networks,” in Proceedings of the IEEE 27th International Conference on Advanced Information Networking and Applications (AINA ’13), pp. 197–204, 2013.
[9] K. Itoh, M. Horiguchi, and T. Kawahara, “Ultra-low voltage nano-scale embedded RAMs,” in Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS ’06), pp. 25–28, 2006.

[10] M. C. Vuran and I. F. Akyildiz, ”Error control in wireless sensor networks: a cross layer analysis,” IEEE/ACM Transactions on Networking, vol. 17, no. 4, pp. 1186–1199, 2009.

[11] B. Chen, L. Tong, and P. K. Varshney, “Channel-aware distributed detection in wireless sensor networks,” IEEE Signal Processing Magazine, vol. 23, no. 4, pp. 16–26, 2006.

[12] J.-Y. Wu, C.-W. Wu, T.-Y. Wang, and T.-S. Lee, “Channel-aware decision fusion with unknown local sensor detection probability,” IEEE Transactions on Signal Processing, vol. 58, no. 3, pp. 1457–1463, 2010.

[13] O. Ozdemir, R. Niu, and P. K. Varshney, “Channel aware target localization with quantized data in wireless sensor networks,” IEEE Transactions on Signal Processing, vol. 57, no. 3, pp. 1190–1202, 2009.

[14] H. Jeon, J. Choi, H. Lee, and J. Ha, “Channel-aware energy efficient transmission strategies for large wireless sensor networks,” IEEE Signal Processing Letters, vol. 17, no. 7, pp. 643–646, 2010.

[15] B. Zhang, A. Arapostathis, S. Nassif, and M. Orshansky, “Analytical modeling of SRAM dynamic stability,” in Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD ’06), pp. 315–322, San Jose, Calif, USA, 2006.

[16] S. B. Wicker, Error Control Systems for Digital Communication and Storage, Prentice Hall, New York, NY, USA, 1995.

[17] S. Mukhopadhyay, C. Schurgers, D. Panigrahi, and S. Dey, “Model-based techniques for data reliability in wireless sensor networks,” IEEE Transactions on Mobile Computing, vol. 8, no. 4, pp. 528–543, 2009.

[18] M. Zuniga and B. Krishnamachari, “Analyzing the transitional region in low power wireless links,” in Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON ’04), pp. 517–526, 2004.

[19] California Data Exchange Center, California Department of Water Resources, 2010, http://cdec.water.ca.gov.

[20] A. Martinez-Sala, J.-M. Molina-Garcia-Pardo, E. Egea-Lopez, J. Vales-Alonso, L. Juan-Llacer, and J. Garcia-Haro, “An accurate radio channel model for wireless sensor networks simulation,” Journal of Communications and Networks, vol. 7, no. 4, pp. 401–407, 2005.

[21] M. H. Hayes, Statistical Digital Signal Processing and Modeling, Wiley, 1996.

[22] S. Gelper, R. Fried, and C. Croux, “Robust forecasting with exponential and holt-winters smoothing,” Journal of Forecasting, vol. 29, no. 3, pp. 285–300, 2010.
