OPTIMAL DISTRIBUTED ESTIMATION IN CLUSTERED SENSOR NETWORKS

Qingjiang Tian and Edward J. Coyle

Center for Wireless System and Applications (CWSA)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN, USA, 47907-2035
{tianq, coyle}@ecn.purdue.edu

ABSTRACT

In a clustered, multi-hop sensor network, a large number of inexpensive, geographically-distributed sensor nodes each make measurements of a source, quantize them into binary sequences, and transmit them over one or more wireless hops to the clusterhead. When all local measurement data has been gathered by the clusterhead, it fuses them into a final estimate about the source. Two sources of error affect the clusterhead's final estimate: (i) local measurement errors made by the sensor nodes because of noisy measurements or unreliable sensors; and (ii) bit errors affecting each hop on the wireless communication channel. Previous work assumed error-free communication or a single-hop cluster. We propose an optimal estimate that accounts for both of these sources of error. We show that this estimate significantly outperforms schemes that consider only the measurement error noise–both in terms of error counts and mean square error.

1. INTRODUCTION

In one scenario envisioned for wireless sensor networks, many small, inexpensive nodes with sensing, processing and wireless communication capabilities are scattered over a field. They self-organize into a clustered network; sense the environment; take measurements of a certain source; and send these measurements to their clusterheads via multi-hop wireless communication. Each clusterhead (CH) fuses the measurement data from the nodes in its cluster to determine a final estimate for that region of the network.

Figure 1 shows a multi-hop cluster in a sensor network. The CH that collects and fuses the local data of the sensor nodes is shown in the center.

We assume that the phenomenon being measured by the nodes in a single cluster has a correlation length that is larger than the diameter of a single cluster. In the ideal case of noise-free measurements and error-free computation, each node should then be making the same estimate about the phenomenon. The estimates made by the sensor nodes and the CH may, however, be incorrect. We focus on two sources of incorrect decisions: (i) local measurement and local estimation errors by each node in a cluster; and (ii) estimation fusion errors at the CH due to communication errors that corrupt the data transmitted by the nodes in the cluster.

For communication induced errors, we take into account the multi-hop communication strategies used within clusters in clustering algorithms like those in [1,2]. Clearly, measurement data forwarded from nodes at the outer edge of a cluster will travel more hops, and thus be more vulnerable to communication error than those from nodes within one hop of the CH. An appropriate weighting of these measurements should thus depend on the number of communication hops they take.

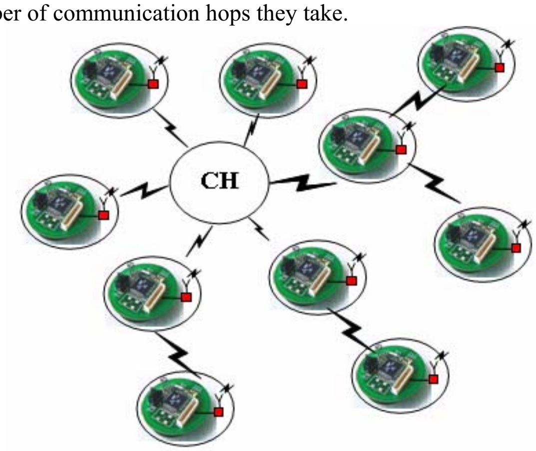

Figure 1. *An example showing MICA2 sensor dots that have self-organized into a two-hop cluster around the CH. Each node has a transmission range that depends on many conditions in its local environment but is generally less than 50 meters.*

In [3], an optimal distributed detection problem has been investigated for a multi-hop cluster. Each node makes a local hard decision, 0/1, and transmits the decision bit to the CH through multiple hops. It was show that the optimal estimate at the CH, based on the Maximum a Posteriori (MAP) criterion, is a weighted median. The error exponent of distributed detection problem with partial noise statistics has been studied in [4]. In [5], an optimal estimate, based on Minimum Mean Square Error (MMSE) criterion, has been proposed -- the estimate weights data from different sensor nodes according to their measurement noise variances. A power scheduling algorithm is also proposed. The asymptotic performance of distributed detection has been studied in [6].

The rest of this paper is organized as follows. The optimal estimate is proposed in Section 2. Some initial numerical results are provided in Section 3. Conclusions and further work are discussed in Section 4.

2. OPTIMAL MAP-BASED ESTIMATION IN A MULTI-HOP CLUSTER

In a K hop cluster, suppose the value of the data source is uniformly distributed over $[0\ U]$. Each node makes one measurement of the source, quantizes the data, encodes the quantized output into an $L-$bit sequence, and transmits the this binary sequence to the processing center. We assume that the quantizer divides the whole interval $[0\ U]$ into $M=2^L$ sub-intervals of equal length $\Delta = U/M$. Without loss of generality, assume that $\Delta=1$. If the measurement data satisfies $x \in (m,(m+1)]$, then the quantizer output will be the binary sequence s_m for decimal number *m*. It is easy to see that each s_m has the same *prior* probability of $1/M$. For this case, the MAP-based estimator of the date is equivalent to the Maximum Likelihood (ML) estimate.

Assume that the communication channel between any pair of sensor nodes and between the CH and any node in the first hop is a Binary Symmetric Channel (BSC). In other words, if s and r are the transmitted and received bits, respectively, then the probability of an error after transmission is given by $p(r=1 \mid s=0)=p(r=0 \mid s=1)=p_c$ and the probability a correct decision is received given by $p(r=0 \mid s=0)=p(r=1 \mid s=1)=1-p_c$. We also assume that each bit of the $L-$bit binary sequence from the same sensor node is equally and independently vulnerable to error. It is easy to show that after a bit is transmitted over k hops that it is received incorrectly with probability:

$$p_{e,k}=\frac{1}{2}-\frac{1}{2}(1-2p_c)^k,\ \mathrm{k}\geq 1. \tag{1}$$

From (1), it is clear that measurements received from nodes in the outer rings of the cluster are less reliable than those from nodes close to the CH. With knowledge of the communication pattern within the cluster, the CH can appropriately weight the data received from each ring and fuse all data into a final estimate about the source.

2.1. Estimation without Measurement Noise

Assume that there is no measurement noise and that all errors are due to quantization error and communication error. Note that quantization error is unavoidable since the number of quantization levels is limited to $M=2^L$. Since the measurement data is uniformly distributed over [0 U], the quantized outputs $s_1, s_2,\ldots,s_M$ each occur with probability $1/M$. For the optimal estimate at the CH, we choose the s_m that maximizes the posterior probability $p(\bar{r} \mid s_m)$, given the received data $\bar{r}$ at the CH.

Definition 1. Assume $X,Y \in \{0,1\}^L$ are two $L-$bit long binary sequences, the Hamming distance between X and Y is defined as $d(X,Y)=\sum_{l=1}^{L}(x_l \oplus y_l)$, where x_l is the l^{th} bit of the vector X.

Theorem 1. Assume that the CH receives $\bar{r}_1=[r_{1,1}, r_{1,2},\ldots,r_{1,N_1}]$ samples from nodes in the first ring, $\bar{r}_2=[r_{2,1}, r_{2,2},\ldots r_{2,N_2}]$ from nodes in the second ring, and so on. Let the cumulated bit error probabilities for samples from different hops are $p_{e,1}, p_{e,2},\ldots,p_{e,K}$. and define $\chi_k=\ln\left((1-p_{e,k})/p_{e,k}\right)$. If $\chi_k s$ can be scaled such that $\chi_1:\chi_2:\ldots:\chi_K=W_1:W_2:\ldots:W_K$, where the $W_k s$ are positive integers and $\gcd(W_1,W_2,\ldots,W_K)=1$. The optimal MAP-based estimate is then:

$$\hat{r}=\underset{s_m, 0\leq m\leq M-1}{\arg\min}\sum_{k=1}^{K}\sum_{n=1}^{N_k}W_k\cdot d(r_{k,n},s_m).$$

Proof: it is easy to show that

$$p(\bar{r}_1,\bar{r}_2,\ldots,\bar{r}_K \mid s_m)=\prod_{k=1}^{K}\prod_{n=1}^{N_k}\left((1-p_{e,k})^{L-d(r_{k,n},s_m)}(p_{e,k})^{d(r_{k,n},s_m)}\right)$$
$$=\left(\prod_{k=1}^{K}\prod_{n=1}^{N_k}\left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n},s_m)}\right)\cdot\prod_{k=1}^{K}(1-p_{e,k})^{N_k\cdot L}.$$

Since $\prod_{k=1}^{K}(1-p_{e,k})^{N_k\cdot L}$ is common to all $p(\bar{r}_1,\bar{r}_2,\ldots,\bar{r}_K \mid s_m)$, maximizing $p(\bar{r}_1,\bar{r}_2,\ldots,\bar{r}_K \mid s_m)$ is equivalent to maximizing $\prod_{k=1}^{K}\prod_{n=1}^{N_k}\left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n},s_m)}$. With the given condition, we know $\ln\left(\frac{1-p_{e,k}}{p_{e,k}}\right)/\ln\left(\frac{1-p_{e,K}}{p_{e,K}}\right)=\frac{W_k}{W_K}$. This leads to:

$$\prod_{k=1}^{K}\prod_{n=1}^{N_k}\left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n},s_m)}$$
$$=\prod_{k=1}^{K}\prod_{n=1}^{N_k}\left((1-p_{e,K})/p_{e,K}\right)^{-W_k\cdot d(r_{k,n},s_m)/W_K} \tag{2}$$
$$=\left(\left((1-p_{e,K})/p_{e,K}\right)^{-\sum_{k=1}^{K}\sum_{n=1}^{N_k}W_k\cdot d(r_{k,n},s_m)}\right)^{1/W_K}.$$

Since $(1-p_{e,K})/p_{e,K}>1$, the maximum of (2) is given by the minimum of the sum of the weighted Hamming distances. Thus the MAP estimate is given by

$$\hat{r}=\underset{s_m, 0\leq m\leq M-1}{\arg\min}\sum_{k=1}^{K}\sum_{n=1}^{N_k}W_k\cdot d(r_{k,n},s_m). \qquad \blacksquare$$

The performance of this estimate can be evaluated via a counting error probability (CEP) and mean square error (MSE). Since there is no measurement noise, suppose the measurement data is correctly quantized into s_m. If the estimate at the CH is different from s_m, i.e., $\hat{r}\neq s_m$, one counting error occurs. The second measure, the MSE, is defined as the mean value of $(s_m-\hat{r})^2$.

In [3], the detection error probability (DEP) is thoroughly investigated in multi-hop cluster sensor networks. Since we assume each bit in one binary sequence is vulnerable to error independently and with identical probability, we can show that, at the CH, one counting error is equivalent to at least one bit error among the $L-$bit binary sequence. Define P_{Bit} as the bit error probability at the CH for each of the L bits. We know that CEP, P_{Count}, is given by

$$P_{Count}=1-(1-P_{Bit})^L. \tag{3}$$

The rightmost bit in a L bit binary sequence has weight 1 while the leftmost bit has weight 2^{L-1}. The MSE resulting from a bit is $\left(2^{l-1}\right)^2$, where l is the position of the bit in the sequence. The estimation MSE at the CH, given P_{Bit}, can thus be derived as

$$\delta_{MSE}=\left(\sum_{l=0}^{L-1}2^{2l}\right)\cdot P_{Bit}. \tag{4}$$

From (3) and (4), we know that

$$\delta_{MSE}=\left(\sum_{l=0}^{L-1}2^{2l}\right)\cdot\left(1-(1-P_{Count})^{1/L}\right) \tag{5}$$

2.2. Estimation with Measurement Noise

Consider the estimation problem in Section 2.1, but now assume that the measurement data is noisy. Assume that the noise is bounded so that the measurement data is still uniformly distributed over $[0\ U]$. We also assume that, due to the measurement noise, with probability $p_{i,j}$, a measurement is quantized into s_j instead of the real value s_i. Define $\overline{p}_i = [p_{i,1} \cdots p_{i,M}]$, note that $\sum_{j=1}^{M} p_{i,j} = 1$.

Theorem 2: In a multi-hop cluster, assume that the CH receives $\overline{r}_1 = [r_{1,1}, r_{1,2}, ..., r_{1,N_1}]$ samples from nodes in the first ring, $\overline{r}_2 = [r_{2,1}, r_{2,2}, ... r_{2,N_2}]$ from nodes in the second ring, and so on. Assume that the bit error probabilities for samples from different rings are $p_{e,1}, p_{e,2}, ..., p_{e,K}$. Define $\chi_k = \ln\left((1-p_{e,k})/p_{e,k}\right)$ and assume that $\chi_k s$ can be scaled up so that $\chi_1 : \chi_2 : ... : \chi_K = W_1 : W_2 : ... : W_K$, where $W_k s$ are positive integers with $\gcd(W_1, W_2, ..., W_K) = 1$. The MAP estimate of the source is given by

$$\hat{r} = \arg\max_{s_m, 0 \le m \le M-1} \sum_{k=1}^{K} \sum_{n=1}^{N_k} \log\left(\sum_{j=1}^{M} p_{i,j} \cdot \left((1-p_{e,K})/p_{e,K}\right)^{-W_k \cdot d(r_{k,n}, s_m)/W_K} \right)$$

Proof: it is easy to show that:

$$\begin{aligned} p(\overline{r}_1, \overline{r}_2, ..., \overline{r}_K \mid s_m) &= \prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot p(r_{k,n} \mid s_j) \right) \\ &= \prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot (1-p_{e,k})^{L-d(r_{k,n}, s_j)} (p_{e,k})^{d(r_{k,n}, s_j)} \right) \\ &= \prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot \left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n}, s_j)} \cdot (1-p_{e,k})^{L} \right) \\ &= \prod_{k=1}^{K} (1-p_{e,k})^{N_k \cdot L} \prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot \left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n}, s_j)} \right) \end{aligned} \quad (6)$$

Since $\prod_{k=1}^{K} (1-p_{e,k})^{N_k \cdot L}$ is common to all $p(\overline{r}_1, \overline{r}_2, ..., \overline{r}_K \mid s_m)$, the MAP estimate is the s_m that maximizes

$$\prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot \left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n}, s_j)} \right).$$

With $\chi_k = \ln\left((1-p_{e,k})/p_{e,k}\right)$ and

$$\chi_1 : \chi_2 : ... : \chi_K = W_1 : W_2 : ... : W_K,$$

it can be shown that

$$\begin{aligned} &\prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot \left((1-p_{e,k})/p_{e,k}\right)^{-d(r_{k,n}, s_j)} \right) \\ &= \prod_{k=1}^{K} \prod_{n=1}^{N_k} \left(\sum_{j=1}^{M} p_{m,j} \cdot \left((1-p_{e,K})/p_{e,K}\right)^{-W_k \cdot d(r_{k,n}, s_j)/W_K} \right). \end{aligned}$$

Since $\log(.)$ is a monotonically-increasing function, the optimal estimate $\hat{r}$ is thus given by

$$\hat{r} = \arg\max_{s_m, 0 \le m \le M-1} \sum_{k=1}^{K} \sum_{n=1}^{N_k} \log\left(\sum_{j=1}^{M} p_{i,j} \cdot \left((1-p_{e,K})/p_{e,K}\right)^{-W_k \cdot d(r_{k,n}, s_m)/W_K} \right)$$

■

Note that Theorem 1 is a special case of Theorem 2 with $p_{i,i} = 1$ and $p_{i,j} = 0$ for $i \ne j$.

3. INITIAL NUMERICAL RESULTS

In this section, we present some initial results on the performance of the proposed weighted estimation algorithm. For now, we only consider the case in which there is no measurement noise, as in Theorem 1. One simple estimation scheme, if we do not consider the effect of multi-hop communication error, is to choose an $\hat{r}$ such that $\hat{r} = \arg\min_{s_m, 0 \le m \le M-1} \sum_{k=1}^{K} \sum_{n=1}^{N_k} d(r_{k,n}, s_m)$. We compared the weighted scheme against this simple unweighted one. The estimates' performance is compared in terms of both CEP and MSE. Assume that each sensor node will transmit a binary sequence of length $L = 5$. Without measurement noise, the quantizer output, s_m, is one of $M = 2^L = 32$ binary sequences with equal probability 1/32. The numerical results are averaged over 10^6 runs.

Figure 2 provides a performance comparison between the weighted estimation scheme and the simple unweighted scheme for different one hop communication error rates p_c. The cluster is assumed to be a 4-hop cluster and there are $(2k-1)$ nodes in ring k. From the Figures we can see that the weighted estimation scheme always outperforms the unweighted scheme, in term of both CEP and MSE. In Fig..b, we also compare the MSEs from both simulations and analytical results derived from CEP. It shows that the MSE is closely related to CEP, as shown in (5).

In Fig. 3, performance comparisons are conducted for different cluster sizes. Assume there are $(2k-1)$ nodes in ring k and the one-hop bit error rate is fixed at $p_c = 0.1$. For a one-hop cluster, the weighted estimation scheme degrades to the simple scheme; this is confirmed in Fig. 3. For multi-hop clusters, we can see that the weighted estimation scheme significantly outperforms the simple one. As the cluster size increases, more samples reach the CH through multiple hops; the accumulated bit errors results in these samples make them more unreliable. These unreliable samples dominate all other samples within the cluster, causing the estimation error to increase after a certain point. On the other hand, the weighted scheme makes proper use of these samples from the outer rings to improve the system performance.

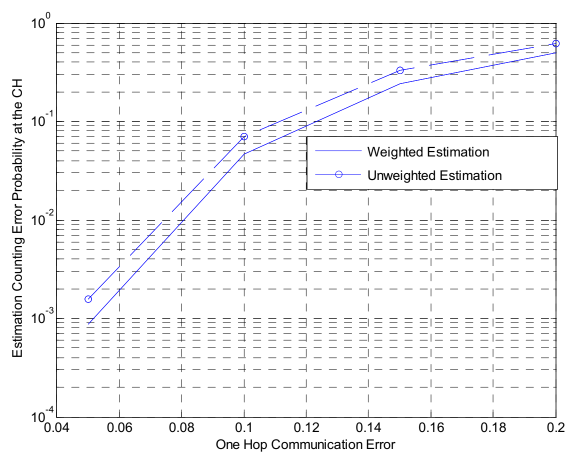

Fig.2.a. Performance comparison in term of CEP.

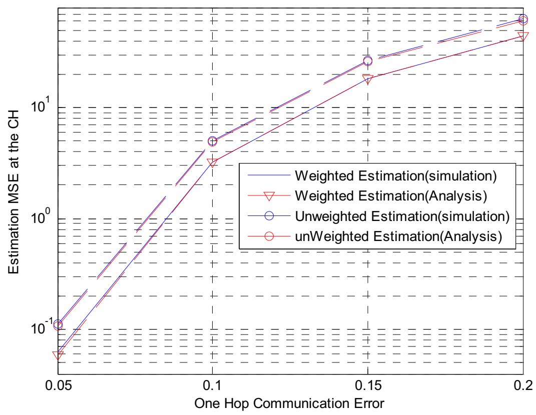

Fig.2.b. Performance comparison in term of MSE.

Fig. 2. Estimation performance comparisons with different p_cs. *The cluster size* $K = 4$ *with* $(2k-1)$ *nodes in ring* k.

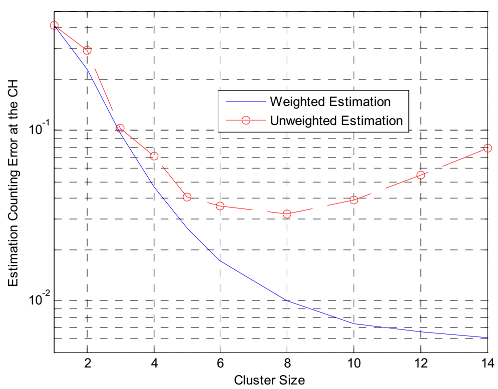

Fig.3.a. Performance comparison in terms of CEP.

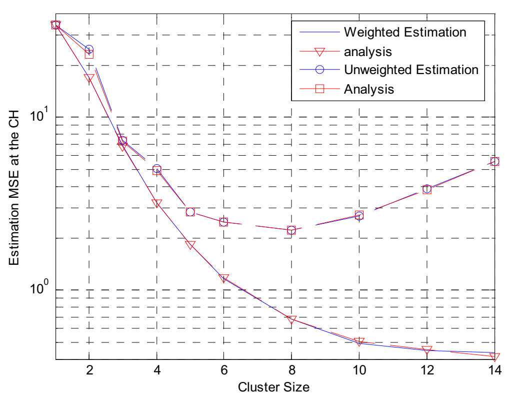

Fig.3.b. Performance comparison in terms of MSE.

Fig. 3. Performance comparisons with different cluster sizes. $p_e = 0.1$, *and there are* $(2k-1)$ *nodes in ring* k.

4. CONCLUSION

Distributed estimation in clustered sensor networks is investigated and the optimal estimate is the quantizer output that minimizes a weighted distance metric. Numerical results showed that the optimal estimate significantly outperforms unweighted schemes. Future work will include investigating the estimation error bound and power scheduling for distributed estimation in sensor networks.

5. REFERENCES

[1] S. Bandyopadhyay and E. J. Coyle, “Minimizing Communication Costs in Hierarchically clustered Networks of Wireless Sensors,” *Computer Networks*, Vol. 44, Issue 1, pp. 1-16, January 2004.

[2] S. J. Baek, G. D. Veciana and X. Su, “ Minimizing Energy Consumption In Large-scale Sensor Networks Through Distributed Data Compression And Hierachical Aggregation,” *IEEE Journal on Selected Areas in Communications*, Vol. 22, No. 6, pp. 1130-1140, Aug. 2004.

[3] Q. J. Tian and Edward J. Coyle, “Optimal Distributed Detection in Clustered Wireless Sensor Networks: The Weighted Median,” Submitted to IEEE INFOCOM 2006.

[4] Z. Q. Luo, “Universal Decentralized Detection in a Bandwidth Constrained Sensor Network,” *IEEE Trans. Signal Processing*, 2004.

[5] J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, “Joint Estimation in Sensor Networks under Energy Constraint,” *IEEE Trans. Signal Processing*, to appear.

[6] J.-F. Chamberland and V. V. Veeravalli, “Asymptotic results for decentralized detection in power constrained wireless sensor networks,” *IEEE J. Select. Areas Commun.*, Vol. 22, No. 6, pp. 1007–1015, Aug. 2004.