HOW DENSE SHOULD A SENSOR NETWORK BE FOR DETECTION APPLICATIONS?

Jean-François Chamberland

Texas Engineering Experiment Station
Dept of Electrical Engineering
Texas A&M University

Venugopal V. Veeravalli

Coordinated Science Laboratory
Dept of Electrical and Computer Engineering
University of Illinois

ABSTRACT

A binary decentralized detection problem is studied in which a collection of wireless sensor nodes provides relevant information about their environment to a fusion center. The observations at the nodes are samples of a finite state Markov process under each hypothesis. The nodes transmit their data to a fusion center over a multiple access channel. Upon reception of the information, the fusion center selects one of the two possible hypotheses. It is assumed that the sensor system is constrained by the capacity of the multiple access channel over which the sensor nodes are transmitting. Thus, as the node density increases, the sensor observations get more correlated, and, furthermore, fewer bits can be transmitted by each sensor node. A framework is presented in this paper for deriving design guidelines relating sensor density to system performance under a total communication constraint. The framework is based on large deviation theory applied to the asymptotic regime where the number of sensor nodes is large. This framework is applied to a specific example to compare the gains offered by having a higher node density with the benefits of getting detailed information from each sensor.

1. STATEMENT OF THE PROBLEM

Wireless sensor systems are often subject to strict power constraints, and wireless sensor nodes must operate on small energy budgets [1]. Accordingly, we seek to better understand the interplay between resource allocation and overall performance in such sensor systems.

In the system that we consider, the sensors are assumed to communicate to the fusion center on a multiple access channel. The achievable rate region for this channel with an appropriate encoding scheme may depend on bandwidth, power, noise density, and maximum tolerable bit error rate at the output of the decoder. In this work, we disregard the specifics of these parameters and assume a constraint on the sum rate from the sensors to the fusion center. Furthermore, we neglect possible communication errors in the transmitted bits. In other words, we assume that the sensor nodes can transmit reliably at a maximum sum rate of R bits per channel use.

To allow a fair comparison between competing designs, we assume that the wireless resources available are identical regardless of implementation specifics, and therefore that all of these systems are subject to the same capacity constraint R. It is therefore of interest to explore how to best allocate these R bits among sensor nodes. More specifically, the decentralized detection design problem consists of selecting integers n and D, where n represents the total number of sensor nodes in the system and D is the number of admissible messages per sensor node, to minimize the probability of error at the fusion center subject to the capacity constraint

$$\sum_{\ell=1}^{n} \lceil \log_2 (D) \rceil \leq R. \qquad (1)$$

This detection framework where a wireless sensor network is subject to a total capacity constraint was first introduced by us in [2]. This framework can be employed to show that, for conditionally independent and identically distributed observations, the gain offered by having more sensors often surpasses the benefits of getting detailed information from each sensor node. Evidence supporting this assertion in a more encompassing setting can also be found in subsequent papers [3, 4].

Most results on the topic of decentralized detection assume that the observations are conditionally independent across sensors. Much less is known about the more general setting where the observations are conditionally dependent. Yet, as sensor nodes are packed more densely in a finite region, it is reasonable to expect their observations to become increasingly correlated. We therefore expand our study of decentralized detection with a total capacity constraint to the scenario where the dependence among observations increases with sensor density.

The scenario we consider is one where the sensors sam-

This research was supported by the National Science Foundation (NSF) under the CAREER/PECASE award CCR-0049089.

ple a spatial stochastic process $\mathbb{X}(s)$ on a line, where s represents the position on the line. The stochastic process can take on one to two possible forms $\mathbb{X}_0(s)$ and $\mathbb{X}_1(s)$, corresponding to the hypotheses H_0 and H_1, respectively. The processes $\mathbb{X}_0(s)$ and $\mathbb{X}_1(s)$ are assumed to be *finite-state* stationary and ergodic Markov processes, with generator matrices $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$, respectively. Denote the finite state space by Σ. Without loss of generality, this space can be identified with the set $\{1, 2, \ldots, |\Sigma|\}$. The stationary probability distribution of the states under hypothesis H_j is denoted by $\pi_j(x)$, $x \in \Sigma$, which satisfies the conditions $[\pi_j(1) \ldots \pi_j(|\Sigma|)]\boldsymbol{Q}_j = 0$ and $\sum_{x\in\Sigma} \pi_j(x) = 1$.

Denote the position of the active sensor nodes by $0 \leq d_1 < \ldots < d_n < \infty$. For convenience we assume that the sensor nodes are equally spaced with $d_k = d(k-1)$, for $k = 1, \ldots, n$ and $d > 0$. The sampled random variable observed at each sensor node under the two hypotheses is then given by

$$H_0 : X_k = \mathbb{X}_0(d_k) = \mathbb{X}_0(d(k-1)), \quad k = 1, \ldots, n$$
$$H_1 : X_k = \mathbb{X}_1(d_k) = \mathbb{X}_1(d(k-1)), \quad k = 1, \ldots, n.$$

Note that, under hypothesis H_j, the sequence of observations $\{X_k\}$ forms a discrete Markov chain with probability transition matrix

$$\boldsymbol{P}_j = e^{\boldsymbol{Q}_j d}.$$

Let $\boldsymbol{P}_j = \{p_j(\ell, m)\}_{\ell,m\in\Sigma}$ be the stochastic matrix described above. The Markov probability measure P_j^σ associated with the transition probability matrix $\boldsymbol{P}_j$ and with the initial state $\sigma \in \Sigma$ can be written as

$$P_j^\sigma(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p_j(\sigma, x_1) \prod_{i=1}^{n-1} p_j(x_i, x_{i+1}).$$

Sensor node k computes a summary $U_k = \gamma(X_k)$ of its own observation, and sends it to the fusion center. As mentioned before, the information is sent over a multiple access channel and it is assumed to be conveyed reliably. Upon reception of the data, the fusion center selects one of the two possible hypotheses.

The design goal is to find an admissible strategy that minimizes the probability of decision error at the fusion center. For rate R, an admissible strategy consists of an integer n denoting the number of sensor nodes, and a compression rule $\gamma : \Sigma \to \{1, 2, \ldots, D\}$ such that (1) holds. For fixed sensor compression rule γ, it is well-known that the class of likelihood ratio tests in which the normalized log-likelihood ratio of the sensor outputs is compared to a threshold is optimal at the fusion center [5]. When the sensor observations are conditionally independent, the normalized log-likelihood ratio reduces to

$$\mathcal{L}(\boldsymbol{u}) = \frac{1}{n}\sum_{k=1}^{n} \log \frac{P_1(u_k)}{P_0(u_k)} = \frac{1}{n}\sum_{k=1}^{n} f(u_k), \qquad (2)$$

where $f(\cdot)$ is given by

$$f(u) \triangleq \log\left(\frac{\sum_{x\in\gamma^{-1}(u)} \pi_1(x)}{\sum_{x\in\gamma^{-1}(u)} \pi_0(x)}\right).$$

For the more general situation where observations are conditionally dependent, the normalized log-likelihood ratio at the fusion center does not admit such a simple form. In these cases, it is usually hard to assess overall performance and to find an optimal system configuration. We circumvent this difficulty by restricting the form of the decision rule at the fusion center. Throughout, we assume that the fusion center makes decisions using a threshold rule on

$$L_n = \frac{1}{n}\sum_{k=1}^{n} f(U_k). \qquad (3)$$

It should be pointed out that this decision rule is not necessarily optimal when the observations across sensor nodes become increasingly correlated. However, it should perform well when the observations are only mildly correlated. Our goal is twofold: (i) we want to characterize overall performance as a function of correlation among observations, and (ii) we wish to study how correlation affects sensor density.

In our work, we are primarily interested in asymptotic regimes where the number of sensor nodes and the area covered by these nodes go to infinity. Asymptotic analysis where the number of nodes becomes large is well-suited for sensor networks since some of these networks are envisioned to contain thousands of nodes. For any reasonable system, the probability of error at the fusion center goes to zero exponentially fast as n grows unbounded. It is then natural to compare system designs based on their exponential rate of convergence to zero. As mentioned before, we wish to provide a fair comparison of competing designs based on their needs in terms of wireless resources. This is achieved by considering error exponents as a function of total bit requirement:

$$-\lim_{R\to\infty} \frac{\log P_e(R)}{R}. \qquad (4)$$

We emphasize that the number of nodes increases linearly with R. The exponential scaling in terms of R is instrumental in comparing systems with equal resources. The performance metric of (4) should provide good design guidelines for systems with a large enough rate constraint and coverage area.

2. LARGE SYSTEM ANALYSIS

In this section, we use results from the theory of large deviations to characterize the error exponent of a threshold test on the empirical means

$$L_n = \frac{1}{n}\sum_{k=1}^{n} f(U_k) = \frac{1}{n}\sum_{k=1}^{n} f(\gamma(X_k)).$$

Let $\boldsymbol{P} = \{p(\ell, m)\}_{\ell,m\in\Sigma}$ be the true transition probability matrix of the Markov chain $\{X_k\}$. Associate with every $\lambda \in \mathbf{R}$ a nonnegative matrix $\boldsymbol{P}_\lambda$, whose elements are

$$p_\lambda(\ell, m) = p(\ell, m)e^{\lambda f(\gamma(\ell))}, \quad \ell, m \in \Sigma. \tag{5}$$

Note that $\boldsymbol{P}$ being irreducible implies that $\boldsymbol{P}_\lambda$ is also irreducible. For each $\lambda \in \mathbb{R}$, let $\rho(\boldsymbol{P}_\lambda)$ denote the Perron-Frobenius eigenvalue of the matrix $\boldsymbol{P}_\lambda$ (see, e.g., [6]). The error exponent of a threshold test on L_n is characterized by the following theorem, whose proof can be found in Dembo and Zeitouni [7].

Theorem 1. *For every $z \in \mathbb{R}$, define*

$$I(z) = \sup_{\lambda\in\mathbb{R}} \{\lambda z - \log \rho(\boldsymbol{P}_\lambda)\}. \tag{6}$$

where $\boldsymbol{P}_\lambda$ is as defined in (5). *Then, the empirical mean L_n satisfies the large deviation principle with the convex, good rate function $I(\cdot)$. Explicitly, for any set $\Gamma \subseteq \mathbf{R}$, and any initial state σ,*

$$\begin{aligned} -\inf_{z\in\Gamma^o} I(z) &\le \liminf_{n\to\infty} \frac{1}{n} \log P^\sigma(L_n \in \Gamma) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log P^\sigma(L_n \in \Gamma) \le -\inf_{z\in\bar{\Gamma}} I(z). \end{aligned}$$

where Γ^o and $\bar{\Gamma}$ denote the interior and closure of Γ, respectively.

3. DENSITY ANALYSIS

Theorem 1 can be employed to assess the performance loss due to dependence across observations, and to study how correlation impacts optimal sensor density. We illustrate this with a simple example. The same procedure can be applied to more complex systems. We study the basic scenario where the observed stochastic process consists of one of two possible Markov signals, with common support $|\Sigma| = 4$. We choose the two generator matrices to be

$$\boldsymbol{Q}_0 = \begin{bmatrix} -q & q & 0 & 0 \\ rq & -(1+r)q & q & 0 \\ 0 & rq & -(1+r)q & q \\ 0 & 0 & rq & -rq \end{bmatrix}$$

and

$$\boldsymbol{Q}_1 = \begin{bmatrix} -rq & rq & 0 & 0 \\ q & -(1+r)q & rq & 0 \\ 0 & q & -(1+r)q & rq \\ 0 & 0 & q & -q \end{bmatrix},$$

where $r \ge 1$. In some sense, this is the simplest Markov scenario where the effects of correlation and density can be analyzed. Note that the two stochastic signals become harder to distinguish as $r \to 1$, while dependence among observations decreases as $q \to \infty$. There are nine possible compression rules, out of which we examine two:

$$\gamma_1(X) = \begin{cases} 1, & X \in \{1, 2\} \\ 2, & X \in \{3, 4\} \end{cases} \quad \text{and} \quad \gamma_2(X) = X.$$

We compare the two corresponding systems by letting the total area covered by the network go to infinity, with the understanding that the amount of wireless resources available per unit area is fixed. In other words, the rate R and the system coverage grow together with their ratio kept constant. This analysis should yield meaningful guidelines for the allocation of system resources in large sensor systems. In particular, these guidelines provide an educated first guess for the design of a large system where the area and the system resources are prespecified.

Since sensor nodes using function $\gamma_1(\cdot)$ transmit one bit of information per channel access while nodes using function $\gamma_2(\cdot)$ send two bit of data, a system subject to a total rate constraint can potentially comprise twice as many nodes of type γ_1. Specifically, for a total rate constraint R, a system may either employ R nodes of type γ_1, or $\left\lfloor \frac{R}{2} \right\rfloor$ nodes of type γ_2. Because of the symmetry in the problem, we gather that the best decision threshold is zero for the two systems. Furthermore, the corresponding error exponents are captured by the value of the good rate function $I(\cdot)$ evaluated at zero.

For a sequence of systems, each with R sensor nodes of type γ_1, the error exponent is given by

$$\begin{aligned} -\lim_{R\to\infty} \frac{\log P_e(R)}{R} &= -\lim_{n\to\infty} \frac{1}{n} \log P_0^\sigma(L_n > 0) \\ &= -\lim_{n\to\infty} \frac{1}{n} \log P_1^\sigma(L_n < 0) = I_1(0) \end{aligned}$$

where the good rate function $I_1(\cdot)$ is given by (6) with $\boldsymbol{P} = e^{\boldsymbol{Q}_0 d}$ and $\boldsymbol{P}_\lambda$ defined by

$$p_\lambda(\ell, m) = p(\ell, m)e^{\lambda f(\gamma_1(\ell))}, \quad \ell, m \in \Sigma.$$

Similarly, for a sequence of systems, each with $\left\lfloor \frac{R}{2} \right\rfloor$ quaternary sensor nodes of type γ_2, the error exponent is given by

$$-\lim_{R\to\infty} \frac{\log P_e(R)}{R} = \frac{1}{2} I_2(0)$$

where the good rate function $I_2(\cdot)$ is also given by (6) with $\boldsymbol{P} = e^{\boldsymbol{Q}_0 2d}$ and

$$p_\lambda(\ell, m) = p(\ell, m)e^{\lambda f(\gamma_2(\ell))}, \quad \ell, m \in \Sigma.$$

Since fewer sensor nodes are employed in the latter sequence of systems, the distance between adjacent nodes is twice as large.

Figure 1 shows the performance of the two systems in terms of error exponent. Note that the Chernoff bound corresponding to conditionally independent observations also

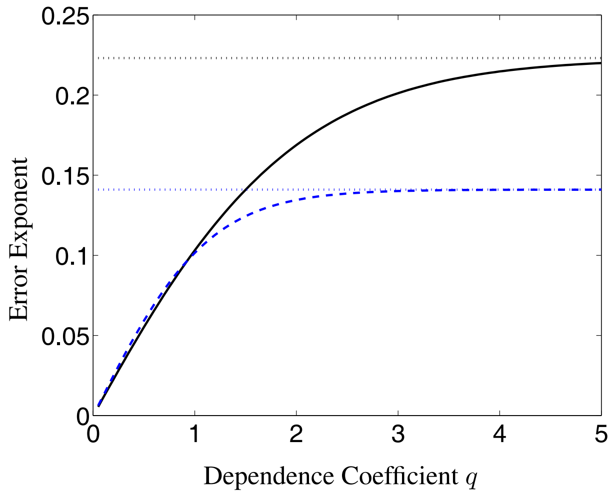

Fig. 1. Error exponent as a function of the dependence coefficient q, with parameters $r = 2$ and $d = 1$. The solid line corresponds to function γ_1, while the dashed line describes performance under γ_2. Note that the distance between adjacent nodes is 1 for system 1, and 2 for system 2.

appears on the graphs for purpose of comparison (the horizontal dotted lines). In this figure, the correlation among observation is varied by changing the value of q in $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$. Not surprisingly, performance decreases as correlation increases. However, it is interesting to note that having more sensor nodes with each node sending fewer bits of information performs very well irrespectively of the dependence coefficient q. When observations are strongly correlated, the two schemes corresponding to γ_1 and γ_2 essentially provides identical performance. On the other hand, as adjacent observations become only weakly dependent, using binary sensor nodes is significantly better.

4. CONCLUSIONS AND DISCUSSION

We considered a decentralized detection problem in which a network of wireless sensors provides relevant information about the state of nature to a fusion center. We addressed the specific case where the sensor observations form a Markov chain under both hypotheses and the sensor network is constrained by the capacity of the multiple access channel over which the wireless sensors are transmitting. We showed through an example that the gain offered by having more nodes often outperforms the benefits of getting detailed information from each sensor even if the sensor observations get more correlated as the density increases. The framework introduced in this paper can be employed to characterize the performance of more complex systems. The finite observation space assumption is not a limitation for practical systems, given that sensor precision is limited. The Markov assumption one the other hand is more restrictive, yet it is much more encompassing than the traditional conditional independence assumption.

5. REFERENCES

[1] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-aware wireless microsensor networks," *IEEE Signal Processsing Magazine*, vol. 19, no. 2, pp. 40–50, March 2002.

[2] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor network," *IEEE Trans. on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.

[3] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, August 2004.

[4] J.-F. Chamberland and V. V. Veeravalli, "The impact of fading on decentralized detection in power constrained wireless sensor networks," in *IEEE ICASSP, Proceedings*, Montreal, Canada, May 2004.

[5] H. V. Poor, *An Introduction to Signal Detection and Estimation,* 2nd edition. New York: Springer, 1994.

[6] R. A. Horn and C. R. Johnson, *Matrix Aalysis*. New York: Cambridge University Press, 1985.

[7] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications,* 2nd edition. New York: Springer, 1998.

[8] 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, August 2002.

[9] R. K. Bahr and J. A. Bucklew, "Optimal sampling schemes for the gaussian hypothesis testing problem," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1677–1686, October 1990.

[10] G. R. Benitz, "Asymptotic results for maximum likelihood estimation with an array of sensors," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1374–1385, July 1993.

[11] W. Bryc and A. Dembo, "On large deviations of empirical measures for stationary gaussian processes," *Stochastic Processes and Their Applications*, vol. 58, pp. 23–34, 1995.