On the Impact of Node Failures and Unreliable Communications in Dense Sensor Networks

Wee Peng Tay, John N. Tsitsiklis, *Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

***Abstract*—We consider the problem of decentralized detection in failure-prone tree networks with bounded height. Specifically, we study and contrast the impact on the detection performance of either node failures (modeled by a Galton–Watson branching process) or unreliable communications (modeled by binary symmetric channels). In both cases, we focus on "dense" networks, in which we let the degree of every node (other than the leaves) become large, and we characterize the asymptotically optimal detection performance. We develop simple strategies that nearly achieve the optimal performance, and compare the performance of the two types of networks.**

***Index Terms*—Decentralized detection, error exponent, sensor networks.**

I. Introduction

We STUDY the performance of sensor networks in the context of decentralized detection. In a typical *parallel* configuration, every sensor makes an observation, summarizes it (e.g., by quantizing it), and sends it to a fusion center, which decides between two or more hypotheses (see, e.g., [1]–[6]). However, in a large scale sensor network, having every sensor send a message directly to the fusion center can be inefficient. Nodes located far away from the fusion center have to expend more energy to transmit their messages reliably, resulting in a shorter lifetime, compared to nodes close to the fusion center. For this reason, there has been considerable interest in more energy efficient configurations such as tree architectures [7]–[14].

In earlier work [15], we studied the detection performance of bounded height tree networks, as the number of nodes increases. For a Neyman–Pearson binary hypothesis testing problem, in which nodes make independent and identically distributed (i.i.d.) observations, we have shown that, under certain mild conditions, the asymptotically optimal detection performance (in terms of the error exponent) is the same as for the parallel configuration. However, in [15], we have not accounted for the possibility of node failures and we have assumed that all messages are received reliably. In this paper, we address these two issues, in the context of "dense" sensor networks. We also aim to obtain qualitative insights into the management of sensor networks.

We model the case of node failures by allowing the number of nodes that transmit messages to a particular node be a random variable with a known distribution. Then, we let the mean of this distribution become asymptotically large, to model a dense network. Parallel configurations with a random number of nodes have been studied by [16]–[18]. In [16] and [17], the authors consider spatially correlated signals and analyzed the performance of a simple but suboptimal strategy. In [18], nodes are assumed to make i.i.d. observations under either hypothesis, quantize their observations using the same quantizer, and use a multiple access protocol that combines the sensor messages in an additive fashion. This paper differs from the previous works in several ways, including the following: 1) we are interested in evaluating the asymptotically optimal detection performance, and in designing asymptotically optimal transmission strategies and 2) we focus on trees with height greater than one. Our results show that for a dense network whose expected number of leaves is n, and under a particular assumption on the distribution of the degree of each node, the asymptotic performance is the same as for a parallel configuration with n leaves, thus establishing that the randomness in the network topology does not lead to performance deterioration.

For the case of unreliable communications, we assume that all nodes are constrained to sending one-bit messages over a binary symmetric channel (BSC) with known crossover probability. To model a dense network, we let the degree of each nonleaf node grow asymptotically large. The case of the parallel configuration is covered by results in [19]. Parallel configurations with a fixed number of nodes and with non-ideal channels between the nodes and the fusion center, have also been studied in [20]–[23]. In this work, we study the effect of unreliable communications on the detection performance of a tree network of height greater than one, and characterize the optimal error exponent. In particular, we show that it is no longer possible to achieve the performance of a parallel configuration, in contrast to the results in [15]. We also consider a scheme that allows a tree network to achieve the same performance as that of a network with reliable communications, but at the expense of increased transmission power. We compare the energy efficiency of such a scheme with that of a parallel configuration and establish that a tree network is preferable.

Finally, we consider the Bayesian version of the problems we have described above, under some additional simplifying assumptions, and characterize the optimal error exponent.

The rest of this paper is organized as follows. In Section II, we state some of the required assumptions and notation. We consider the case of node failures in Section III and the case of unreliable communications in Section IV, both under a

Manuscript received July 5, 2007; revised November 12, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Danilo P. Madic. This research was supported in part by the National Science Foundation under Contracts ECCS-0701623, ECS-0426453, ANI-0335256, and ECS-0636519 and by DoCoMo USA Labs. A preliminary version of this paper was presented at the IEEE Wireless Communications and Networking Conference, Hong Kong, March, 2007.

The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: wptay@mit.edu; jnt@mit.edu; moewin@mit.edu).

Digital Object Identifier 10.1109/TSP.2007.914343

Neyman–Pearson formulation. In Section V, we consider the impact of node failures and unreliable communications in a Bayesian setting. In Section VI, we conclude and summarize this paper.

II. Preliminaries

In this section, we describe the basic model and introduce some notation and assumptions. Most elements of the basic model are borrowed from [15], which we reproduce in summary form as follows.

A. Tree Architectures

We consider a *tree network*, modeled by a directed tree $T=(V,E)$, where V is the set of nodes and E the set of directed arcs. One of the nodes (the "root" of the tree), denoted by f, is designated as the *fusion center* and we assume that all arcs are oriented so that they point towards the fusion center. We say that node u is a *predecessor* of node v if there exists a directed path from u to v. We also say that v is a *successor* of u. An immediate predecessor of node v is a node u such that $(u,v)\in E$. An immediate successor is similarly defined. Let the set of immediate predecessors of v be $C(v)$. Let $l(v)$ be the number of leaves in the sub-tree rooted at node v. With this definition, $l(f)$ is the total number of leaves.

The *length* of a path is the number of arcs in the path. The *height* of a tree is the length of a longest path from a leaf to the root. Suppose that T has height h. If a node is k hops away from the fusion center, it is said to be at *level* $h-k$. Hence, the fusion center is at level h and its immediate predecessors (nodes in the set $C(f)$) are at level $h-1$. Furthermore, any node at level 0 is a leaf.

B. Strategies

We are given a probability space and two probability measures $\mathbb{P}_0$ and $\mathbb{P}_1$ on this space, which correspond to two hypotheses H_0 and H_1. We will use $\mathbb{E}_j$ and var_j to denote expectation and variance with respect to $\mathbb{P}_j$. Under hypothesis H_j ($j=0,1$), every leaf v observes an i.i.d. random variable X_v, which takes values in a set $\mathcal{X}$ and has distribution $\mathbb{P}_j^X$. It then summarizes its observation X_v using a *transmission function* γ_v and transmits a message $Y_v=\gamma_v(X_v)$ to its immediate successor. For simplicity, we constrain the messages to be symbols in a fixed transmission alphabet $\mathcal{T}$, so that γ_v maps $\mathcal{X}$ to $\mathcal{T}$. In general, the transmission alphabet $\mathcal{T}$ has a smaller cardinality than $\mathcal{X}$, e.g., $\mathcal{X}$ could be the set of real numbers $\mathbb{R}$, while $\mathcal{T}$ could be a finite set. In Sections IV and V, we will explicitly assume that $\mathcal{T}=\{0,1\}$. Let Γ be a given set of transmission functions that a leaf can choose from.

A nonleaf node v uses a transmission function γ_v to encode and transmit a summary $Y_v=\gamma_v(\{Y_u : u\in C(v)\})$ of its received messages to its immediate successor. Suppose that the number $|C(v)|$ of immediate predecessors of v is $d\geq 1$. Then, the transmission function γ_v maps $\mathcal{T}^d$ to $\mathcal{T}$.[1] Similarly, the root f uses a *fusion rule* γ_f, which depends on the received messages to make a decision. (The fusion rule γ_f can be regarded as a "transmission function" for the node f.) Let Y_f be a binary value random variable indicating the decision of the fusion center.

A *strategy* (for the tree T) is a collection of transmission functions, one for each node, and a fusion rule.[2] Throughout this paper, we assume that there is a multiple access protocol in place, so that every node can distinguish the messages it receives from each of its immediate predecessors. In particular, transmissions from one node do not interfere with those of another node. This can for example, be a random access protocol or a time/frequency division multiplexing scheme.

C. Assumptions and Notations

For any $\gamma\in\Gamma$, let $\mathbb{P}_j^\gamma$ be the distribution of the random variable $Y=\gamma(X)$, where X has distribution $\mathbb{P}_j^X$. We quantify the information content of Y in terms of the Kullback–Leibler (KL) divergence, defined by (recall that $\mathbb{E}_0$ is the expectation operator under the probability measure $\mathbb{P}_0$)

$$\mathrm{D}\left(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma\right)=\mathbb{E}_0\left[\log\frac{\mathrm{d}\mathbb{P}_0^\gamma}{\mathrm{d}\mathbb{P}_1^\gamma}\right].$$

Note that $\mathrm{D}(\mathbb{P}_0^\gamma\|\mathbb{P}_1^\gamma)\geq 0$, with the inequality being strict as long as the measures $\mathbb{P}_0^\gamma$ and $\mathbb{P}_1^\gamma$ are not indistinguishable. The following assumptions are standard, and are the same as those made in [15].

Assumption 1: The measures $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent, i.e., they are absolutely continuous with respect to (w.r.t.) each other. Furthermore, there exists some $\gamma\in\Gamma$ such that $\mathrm{D}(\mathbb{P}_0^\gamma\|\mathbb{P}_1^\gamma)>0$.

Assumption 2: $\mathbb{E}_0[\log^2(\mathrm{d}\mathbb{P}_1^X/\mathrm{d}\mathbb{P}_0^X)]<\infty$.

III. Node Failures

We model node failures by letting the number of immediate predecessors of each node be random variables with known distributions. Although [16]–[18] have studied variations of this problem in a different context, they specifically assumed a Poisson distribution and considered only the parallel configuration. Our formulation involves trees with a general height $h\geq 1$ and distributions from a somewhat larger family. The main reason for introducing this larger family of distributions is to facilitate comparison with the results in Section IV.

Let h be a positive integer. We form a random tree according to a Galton–Watson branching process [24] with h stages. Consider the fusion center f. Let $N_f=|C(f)|$ be a nonnegative integer random variable, with marginal law μ_h. For each node v in the random set $C(f)$, we let $N_v=|C(v)|$ be i.i.d. random variables with distribution μ_{h-1}. We continue this process until the level 0 nodes are reached. Hence, each level k node v (with $k\geq 1$) has N_v immediate predecessors, where N_v is a random variable with law μ_k. Furthermore, we assume that all these random variables are independent, and are also independent of the hypothesis. We call such a tree a *GW tree* of height h. We will sometimes use M_k to denote a generic random variable with law μ_k.

[1] In a variant of the model, we could let each nonleaf node v obtain an additional independent observation X_v. However, we will be focusing on the asymptotic regime where nonleaf nodes typically have a large number of predecessors. In this regime, such additional observations will not improve the asymptotic detection performance, and it can be shown that there is no loss of generality if these observations are omitted from the model.

[2] In general, the transmission functions for each node, and the fusion rule can be randomized functions, but we have avoided any discussion of randomization for simplicity, and because randomization does not improve the asymptotically optimal performance.

Let $\lambda_k = \mathbb{E}[M_k] < \infty$ be the mean[3] of the distribution μ_k and let $\lambda^* = \min_{1\le k\le h} \lambda_k$. We consider the case of asymptotically large λ^* to model a dense network, i.e., we let $\lambda_k \to \infty$ for all k and allow the laws μ_k to vary accordingly. Strictly speaking, we are dealing with a sequence of random tree networks: each tree in the sequence corresponds to a different choice of the parameters λ_k and these parameters tend to infinity along this sequence. However, we keep this underlying sequence hidden (and implicit), to prevent overburdening the notation. Let $\lambda(k) = \prod_{i=1}^{k} \lambda_i$, which is the expected number of leaves that are predecessors of a level k node.

We make the following assumption. The assumption is satisfied if M_k has Poisson distribution with mean λ_k or if there is a constant $p \in (0,1)$ such that M_k has a Binomial distribution $\mathcal{B}(n_k, p)$ with mean $\lambda_k = n_k p$. If every N_v has a Binomial distribution, a GW tree can be interpreted as a deterministic tree network with erasure channels between nodes.

Assumption 3: Let M_k be random variables with distribution μ_k and mean λ_k, $k = 1, \ldots, h$. We have

$$\mathrm{var}[M_k]/\lambda_k^2 \to 0, \quad \text{as } \lambda_k \to \infty. \tag{1}$$

Suppose that the distributions $\mu_1, \ldots, \mu_h$ have been fixed. A transmission *policy* for a node v specifies the transmission function of v, for each realization of the in-degree N_v. Similarly, a *GW-strategy* is defined as a mapping, which for any realization of the random tree, specifies a strategy (as defined at the end of Section II-B) for that tree. Note that a GW-strategy requires, in general, global information on the structure of the realized tree and may be hard to implement. Given a GW-strategy π and a set of distributions $\mu = (\mu_1, \ldots, \mu_h)$, let β_π be the resulting Type II error probability $\mathbb{P}_1(Y_f = 0)$ at the fusion center. (This is an average over all possible realizations of the tree, as well as over the distribution of the observations.) Let us fix some $\alpha \in (0,1)$. Let β^* be the infimum of β_π, over all GW-strategies π, subject to the constraint that the Type I error probability $\mathbb{P}_0(Y_f = 1)$ is less than or equal to α. Our goal is to characterize the optimal error exponent[4]

$$\limsup_{\lambda^* \to \infty} \frac{1}{\lambda(h)} \log \beta^*.$$

(Recall that $\lambda(h) = \prod_{k=1}^{h} \lambda_k$ is the expected number of leaves, as determined by μ.)

Given a GW-strategy and a level k node v, let L_v be the log-likelihood ratio (more formally, the logarithm of the Radon–Nikodym derivative) of the distribution of Y_v under H_1 with respect to that under H_0. In particular, if $\mathcal{T}$ is finite, and if $Y_v = y$, then $L_v = \log(\mathbb{P}_1(Y_v = y)/\mathbb{P}_0(Y_v = y))$. Note that if v is a leaf that uses the transmission function γ, then $\mathbb{E}_0[L_v] = -\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma)$. If v is at level $k \ge 1$, we define the log-likelihood ratio of the messages it receives by

$$S_v = \sum_{u \in C(v)} L_u$$

[3]When dealing with the distribution of the GW tree, we will use the notation $\mathbb{P}$, $\mathbb{E}$, and var, since the distribution is the same under either hypothesis.

[4]Note that according to our sign convention, error exponents are negative.

where the sum is taken to be 0 if $C(v)$ is empty.

Motivated by the ϵ-optimal strategies for non-random tree networks [15], we will be interested in the case where nodes v at some level $k \ge 1$ use a transmission policy [called a mean-normalized log-likelihood ratio (MLLR) quantizer] that results in a message Y_v of the form

$$Y_v = \begin{cases} 0, & \text{if } S_v/\lambda(k) \le t \\ 1, & \text{otherwise} \end{cases}$$

for some threshold t. We assume that all nonleaf nodes are allowed to use MLLR quantizers.

For deterministic network topologies, i.e., if $N_v = \lambda_k$ almost surely, for all level k nodes v, our previous work [15] shows that the Type II error probability decays exponentially fast with $\lambda(h)$, at rate g_P^*, where

$$g_P^* = -\sup_{\gamma \in \Gamma} \mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma).$$

(This is the same as the rate of error decay in a parallel configuration, hence the notation g_P^*.) The proposition below shows that this remains true for a GW tree.

Proposition 1: Suppose that Assumptions 1–3 hold and that $\alpha \in (0,1)$. The optimal error exponent of a GW tree of height h is given by

$$\lim_{\lambda^* \to \infty} \frac{1}{\lambda(h)} \log \beta^* = g_P^*. \tag{2}$$

Furthermore, for any $\epsilon \in (0, -g_P^*)$ and any large enough λ^*, the following GW-strategy satisfies the Type I error probability constraint and its error exponent is bounded above by $g_P^* + \epsilon$:

1) each leaf uses the same transmission function $\gamma \in \Gamma$, with $-\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) \le g_P^* + \epsilon/2 < 0$;
2) for $k \ge 1$, every level k node uses a MLLR quantizer with threshold $t_k = -\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) + \epsilon/2^{h-k+1}$.

To prove the proposition, we will first lower bound the optimal error exponent. We will then derive a matching upper bound by showing that the proposed GW-strategy comes within ϵ of the lower bound.

A. Lower Bound

In this section, we show that in the limit, as $\lambda^* \to \infty$ and for any GW-strategy, the error exponent is lower bounded by g_P^*. We will use the following elementary fact, proved in the Appendix.

Lemma 1: Suppose that X and Y are non-negative random variables with $\mathbb{E}[X] \le a$ and $\mathbb{E}[Y] \le b$, and that the event A has probability $\mathbb{P}(A) > c_1 + c_2$, where $c_1, c_2 > 0$. Then, there exists some $\omega \in A$ such that $X(\omega) \le a/c_1$ and $Y(\omega) \le b/c_2$.

In the following lemma, we show that $l(f)$ (the actual number of leaves) and $\lambda(h)$ (the expected number of leaves) are close (in probability), in the limit of large λ^*. The proof is in the Appendix.

Lemma 2:

(a) $\mathbb{E}[l(f)] = \lambda(h)$ and $\mathrm{var}[l(f)]/\lambda^2(h) \to 0$, as $\lambda^* \to \infty$.

(b) For all $\delta > 0$, $\mathbb{P}(|l(f)/\lambda(h) - 1| > \delta) \to 0$, as $\lambda^* \to \infty$.

We are now ready to prove the lower bound for the optimal error exponent.

Lemma 3: Suppose that Assumptions 1–3 hold, and that $\alpha \in (0,1)$. Then

$$\liminf_{\lambda^*\to\infty} \frac{1}{\lambda(h)} \log \beta^* \geq g_P^*.$$

Proof: Suppose that $g := \liminf_{\lambda^*\to\infty}(1/\lambda(h))\log\beta^* < g_P^*$. Fix $\epsilon > 0$ and $\delta \in (0,1)$ such that $(g+\epsilon)/(1+\delta) < g_P^*$. Then, there exists a sequence of distributions $(\mu_1,\ldots,\mu_h)$ along which $\lambda^* \to \infty$, such that for the kth element of that sequence we have $\lambda(h) = \zeta_k$, where $\zeta_1 \geq 1$, $\zeta_{k+1} \geq ((1+\delta)/(1-\delta))\zeta_k$, $k = 1,2,\ldots$, and

$$\lim_{k\to\infty} \frac{1}{\zeta_k} \log \beta^* = g.$$

Let $\mathcal{G}$ be the set of all trees with height less than or equal to h, and let R_k be a random tree, generated according to the GW process. It follows that there exists some $K_1 > 0$ such that for all $k \geq K_1$, we have

$$\begin{aligned} \mathbb{E}\left[\mathbb{P}_1(Y_f = 0|R_k)\right] &= \mathbb{P}_1(Y_f = 0) \leq e^{\zeta_k(g+\epsilon)} \\ \mathbb{E}\left[\mathbb{P}_0(Y_f = 1|R_k)\right] &= \mathbb{P}_0(Y_f = 1) \leq \alpha. \end{aligned}$$

Fix a $c \in (\alpha, 1)$. From Lemma 2(b), $\mathbb{P}(l(f) \in [(1-\delta)\zeta_k, (1+\delta)\zeta_k]) \to 1$, as $k \to \infty$. Since $(1+c)/2 < 1$, we can choose a $K \geq K_1$, such that for all $k \geq K$

$$\mathbb{P}\left(l(f) \in [(1-\delta)\zeta_k, (1+\delta)\zeta_k]\right) > \frac{1+c}{2} = \frac{1-c}{2} + c.$$

Using Lemma 1, for each $k \geq K$, there exists some tree $r_k \in \mathcal{G}$ with n_k leaves, where $n_k \in [(1-\delta)\zeta_k, (1+\delta)\zeta_k]$, so that

$$\mathbb{P}_1(Y_f = 0|R_k = r_k) \leq \frac{2}{1-c} e^{\zeta_k(g+\epsilon)} \tag{3}$$

$$\mathbb{P}_0(Y_f = 1|R_k = r_k) \leq \frac{\alpha}{c} < 1. \tag{4}$$

From (3)

$$\begin{aligned} \frac{1}{n_k} \log \mathbb{P}_1(Y_f = 0|R_k = r_k) &\leq \frac{\zeta_k}{n_k}(g+\epsilon) + \frac{1}{n_k}\log\frac{2}{1-c} \\ &\leq \frac{g+\epsilon}{1+\delta} + \frac{1}{n_k}\log\frac{2}{1-c}. \end{aligned}$$

Letting $k \to \infty$, we obtain

$$\limsup_{k\to\infty} \frac{1}{n_k} \log \mathbb{P}_1(Y_f = 0|R_k = r_k) \leq \frac{g+\epsilon}{1+\delta} < g_P^*.$$

Recall that g_P^* is the optimal Type II error exponent (as $k \to \infty$) of a parallel configuration with n_k nodes sending messages directly to the fusion center, subject to the constraint that the Type I error probability is less than or equal to α/c [cf. (4) and [19]]. Since such a parallel configuration can simulate the tree r_k, we obtain a contradiction, which proves the desired result. ■

B. Achievability

In this subsection, we fix some $\epsilon \in (0, -g_P^*)$, consider a GW-strategy of the form described in Proposition 1, and show that it performs as claimed. In particular, for $k \geq 1$, every level k node v sends a 0 (or, for the fusion center, it declares H_0) iff $S_v \leq \lambda(k)t_k$.

We first show that this strategy results in a Type II error exponent within ϵ of g_P^*. Consider a node v at level $k \geq 1$. Since $\exp(-S_v)$ is the ratio of the likelihood under H_0 to that under H_1, of the received messages at node v, we have $\mathbb{E}_1[\exp(-S_v)] = 1$. Hence, from the Chernoff bound, we obtain

$$\frac{1}{\lambda(k)} \log \mathbb{P}_1\left(\frac{S_v}{\lambda(k)} \leq t_k\right) \leq \frac{1}{\lambda(k)} \log\left(e^{\lambda(k)t_k}\mathbb{E}_1[e^{-S_v}]\right) = t_k. \tag{5}$$

In particular, for $v = f$, we have $k = h$ and

$$\begin{aligned} \frac{1}{\lambda(h)} \log \mathbb{P}_1\left(S_f \leq \lambda(h)t_h\right) \leq t_h &= -\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) + \frac{\epsilon}{2} \\ &\leq g_P^* + \epsilon. \end{aligned} \tag{6}$$

By taking $\epsilon \to 0$ in (6), we obtain the claimed upper bound on the Type II error exponent.

It only remains to verify that this strategy meets the Type I error constraint, when λ^* is sufficiently large. This is accomplished by the following lemma.

Lemma 4: Suppose that Assumptions 1–3 hold. Let v be a level k node, with $k \geq 1$. For the particular GW-strategy proposed in Proposition 1, we have $\mathbb{P}_0(Y_v = 0) \to 1$, as $\lambda^* \to \infty$.

Proof: We proceed by induction on k. We start by considering the case $k = 1$. Let u be a typical immediate predecessor of v. We have

$$\frac{\mathbb{E}_0[S_v]}{\lambda(1)} = \frac{\mathbb{E}[N_v]}{\lambda(1)} \cdot \mathbb{E}_0[L_u] = \mathbb{E}_0[L_u] = -\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma).$$

Furthermore, using a well-known formula for the variance of the sum of a random number of i.i.d. random variables

$$\frac{\mathrm{var}_0[S_v]}{\lambda^2(1)} = \frac{\mathrm{var}[N_v]\left(\mathbb{E}_0[L_u]\right)^2 + \mathbb{E}[N_v]\mathrm{var}_0[L_u]}{\lambda^2(1)}$$

which converges to zero as $\lambda(1) \to \infty$, because $\mathrm{var}[N_v]/\lambda^2(1)$ converges to zero (Assumption 3), $\mathbb{E}_0[L_u] = -\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) < \infty$, $\mathbb{E}[N_u] = \lambda(1)$, and $\mathrm{var}_0[L_u] \leq \mathbb{E}_0[L_u^2] < \infty$ (from Assumption 2 and Proposition 3 of [19]). Since the threshold t_1 used by v satisfies $-\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) < t_1$, Chebychev's inequality yields $\mathbb{P}_0(S_v > \lambda(1)t_1) \to 0$, and, therefore, $\mathbb{P}_0(Y_v = 0) \to 1$.

Suppose now that the induction hypothesis holds for $k-1$, where $k \geq 2$. Let v be a level k node and let u be a typical immediate predecessor of v. Using the facts $\mathbb{P}_0(Y_u = 0) \to 1$

and $\mathbb{P}_0(Y_u = 1) \to 0$ in the second equality that follows, we have

$$\begin{aligned}\limsup_{\lambda^*\to\infty} \frac{\mathbb{E}_0[L_u]}{\lambda(k-1)} &= \limsup_{\lambda^*\to\infty} \frac{1}{\lambda(k-1)} \left(\mathbb{P}_0(Y_u=0)\log\frac{\mathbb{P}_1(Y_u=0)}{\mathbb{P}_0(Y_u=0)}\right.\\ &\qquad\left.+\mathbb{P}_0(Y_u=1)\log\frac{\mathbb{P}_1(Y_u=1)}{\mathbb{P}_0(Y_u=1)}\right)\\ &= \limsup_{\lambda^*\to\infty} \frac{1}{\lambda(k-1)} \log \mathbb{P}_1(Y_u=0)\\ &= \limsup_{\lambda^*\to\infty} \frac{1}{\lambda(k-1)} \log \mathbb{P}_1\left(S_u \le \lambda(k-1)t_{k-1}\right)\\ &\le t_{k-1}\end{aligned}$$

where the last inequality follows from (5) applied to u.

Using a similar argument, we have

$$\begin{aligned}\limsup_{\lambda^*\to\infty} \frac{\mathrm{var}_0[L_u]}{\lambda^2(k-1)} &\le \limsup_{\lambda^*\to\infty} \frac{\mathbb{E}_0\left[L_u^2\right]}{\lambda^2(k-1)}\\ &= \limsup_{\lambda^*\to\infty} \frac{1}{\lambda^2(k-1)}\\ &\quad\times \log^2 \mathbb{P}_1\left(S_u \le \lambda(k-1)t_{k-1}\right)\\ &\le t_{k-1}^2. \end{aligned} \tag{7}$$

We then obtain

$$\begin{aligned}\limsup_{\lambda^*\to\infty} \frac{1}{\lambda(k)}\mathbb{E}_0[S_v] &= \limsup_{\lambda^*\to\infty} \frac{1}{\lambda(k)}\mathbb{E}[N_v]\mathbb{E}_0[L_u]\\ &= \limsup_{\lambda^*\to\infty} \frac{1}{\lambda(k-1)}\mathbb{E}_0[L_u]\\ &\le t_{k-1}.\end{aligned}$$

Furthermore

$$\frac{\mathrm{var}_0[S_v]}{\lambda^2(k)} = \frac{\mathrm{var}[N_v]\left(\mathbb{E}_0[L_u]\right)^2 + \mathbb{E}[N_v]\mathrm{var}_0[L_u]}{\lambda_k^2 \cdot \lambda^2(k-1)}$$

which converges to zero as $\lambda^* \to \infty$, because $\mathrm{var}[N_v]/\lambda_k^2 \to 0$ (Assumption 3), $\mathbb{E}[N_v]/\lambda_k^2 = 1/\lambda_k \to 0$, and both $\mathbb{E}_0[L_u]/\lambda(k-1)$ and $\mathrm{var}_0[L_u]/\lambda^2(k-1)$ are bounded. Since $t_{k-1} < t_k$, Chebychev's inequality shows that $\mathbb{P}_0(S_v > \lambda(k)t_k) \to 0$, and, therefore, $\mathbb{P}_0(Y_v = 0) \to 1$. ■

C. Discussion

We have shown that the optimal error exponent for a tree network with node failures is g_P^*, the same as for a parallel configuration with a large but deterministic number of nodes, and developed a strategy that achieves the optimal performance, as close as desired. In our ϵ-optimal strategy, every nonleaf node uses an MLLR quantizer. Hence, there is no loss in optimality if we restrict each of the nonleaf nodes to sending only one bit.

Another advantage of this strategy is that every nonleaf node only needs to know the received messages from its immediate predecessors and the distributions $\mu_1, \ldots, \mu_h$; no additional information on the topology of the realized tree is required. While it might be possible, in a static network, as part of the setup process, to inform each node of the topology of the network, this would be too difficult or costly in a mobile or time-varying network. The model that we have adopted, i.e., modeling the immediate predecessors of each sensor as a random set, can be applied to a mobile network, in which a node does not know *a priori* how many nodes will be within transmission range. See [16] for a related model, employed in a similar spirit.

IV. Unreliable Communications

In this section, we consider the case where messages are restricted to be binary and the channel between any two nodes is a binary symmetric channel (BSC) with known crossover probability $\eta \in (0, 1/2)$. Let $T_n = (V_n, E_n)$ be a tree with n nodes. The sequence of trees $(T_n)_{n\ge 1}$ models the evolution of the network as more nodes are added. We assume that for some n_0, and for all $n \ge n_0$, T_n is an h-uniform tree [15], i.e., all leaves are exactly h hops away from the fusion center (this is done for simplicity, to reduce the number of cases that we need to consider; an extension to more general types of trees is possible). For every nonleaf node v, we assume that $|C(v)| \ge c_n$, for some sequence c_n of positive integers that diverges to infinity as n increases. Similar to Section III, we are interested in characterizing the Type II error exponent at the fusion center, when the Type I error probability is constrained to be less than or equal to a given $\alpha \in (0, 1)$. However, in this case, it turns out that the relevant error exponent is

$$\limsup_{n\to\infty} \frac{1}{|C(f)|} \log \beta^*$$

where β^* is the minimum Type II error probability at the fusion center, for the tree T_n, optimized over all strategies that satisfy the Type I error constraint.[5] Note that we have normalized the error exponent using $|C(f)|$ (instead of $l(f)$, the total number of leaves), even though every leaf makes an observation. The reason for this will become apparent in Proposition 3 below.

Consider a nonleaf node v. It receives a message from each node $u \in C(v)$, and forms a message Y_v, which it sends to its immediate successor w. Because of the noisy channel, the message received by w, denoted by Z_v, may be different from Y_v. Let $\bar{L}_v$ be the log-likelihood ratio of the distribution of Z_v under H_1 with respect to that under H_0. Since Z_v is binary, the random variable $\bar{L}_v$ takes one of the two values $\log(\mathbb{P}_1(Z_v = z)/\mathbb{P}_0(Z_v = z))$, $z = 0,1$, depending on whether Z_v is 0 or 1. Let

$$S_v = \sum_{u\in C(v)} \bar{L}_u$$

which is the sum of the log-likelihood ratios of the received messages at node v.

We will be interested in the case where nodes v at some level $k \ge 1$ use a transmission policy [called a log-likelihood ratio quantizer (LLRQ)] that results in a message Y_v of the form

$$Y_v = \begin{cases} 0, & \text{if } S_v/|C(v)| \le t \\ 1, & \text{otherwise}\end{cases}$$

for some threshold t.

[5]To simplify notation, we are suppressing the dependence of β^* and $|C(f)|$ on n.

A. Case $h = 1$

Let us first consider the simple case where $h = 1$, i.e., the parallel configuration. For every $\gamma \in \Gamma$, let

$$q_j^\gamma(0) = (1-\eta)\mathbb{P}_j(\gamma(X) = 0) + \eta\mathbb{P}_j(\gamma(X) = 1)$$
$$q_j^\gamma(1) = (1-\eta)\mathbb{P}_j(\gamma(X) = 1) + \eta\mathbb{P}_j(\gamma(X) = 0)$$

and

$$e_{0,\gamma} = q_0^\gamma(0) \cdot \log\frac{q_1^\gamma(0)}{q_0^\gamma(0)} + q_0^\gamma(1) \cdot \log\frac{q_1^\gamma(1)}{q_0^\gamma(1)}$$
$$e_{1,\gamma} = q_1^\gamma(0) \cdot \log\frac{q_1^\gamma(0)}{q_0^\gamma(0)} + q_1^\gamma(1) \cdot \log\frac{q_1^\gamma(1)}{q_0^\gamma(1)}.$$

For an interpretation, note that if u is a leaf that employs the transmission function γ, then $e_{j,\gamma} = \mathbb{E}_j[\bar{L}_u]$. Let $e_0 = \inf_{\gamma\in\Gamma} e_{0,\gamma}$. The following proposition follows immediately from [19].

Proposition 2: Suppose that Assumptions 1 and 2 hold. Then, for $h = 1$, and for any $\alpha \in (0,1)$, we have

$$\lim_{n\to\infty} \frac{1}{|C(f)|} \log\beta^* = e_0.$$

Furthermore, the optimal error exponent does not change if we restrict all the leaves to use the same transmission function $\gamma \in \Gamma$.

As shown in [19], the optimal error exponent e_0 can be achieved to within some ϵ, by letting all leaves use a transmission function γ that satisfies $e_{0,\gamma} \le e_0 + \epsilon/2$, and letting the fusion center use a LLRQ with threshold $t = e_{0,\gamma} + \epsilon/2$.

B. General Case

We henceforth assume that $h \ge 2$. We have the following proposition, which shows that the optimal error exponent is the same as that of a parallel configuration in which the nodes in $C(f)$ have perfect knowledge of the true hypothesis. Intuitively, as n becomes large, each node $v \in C(f)$ discriminates between the two hypotheses with vanishing probabilities of error. Let $\mathrm{Bern}(\eta)$ denote the Bernoulli distribution on {0,1} that takes value 1 with probability η. Let

$$D(\eta) = \eta\log\frac{\eta}{1-\eta} + (1-\eta)\log\frac{1-\eta}{\eta}$$

which is the KL divergence function of $\mathrm{Bern}(1-\eta)$ w.r.t. $\mathrm{Bern}(\eta)$.

Proposition 3: Suppose that Assumptions 1 and 2 hold, $h \ge 2$, and $\alpha \in (0,1)$. Then, the optimal error exponent is

$$\lim_{n\to\infty} \frac{1}{|C(f)|} \log\beta^* = -D(\eta) < 0.$$

Furthermore, for any $\epsilon > 0$, as $n \to \infty$, the following strategy satisfies the Type I error probability constraint, and also satisfies $\limsup_{n\to\infty}(1/|C_n(f)|)\log\mathbb{P}_1(Y_f = 0) \le -D(\eta) + \epsilon$:

1) all leaves use the same transmission function $\gamma \in \Gamma$, where γ is chosen so that $\mathbb{P}_0(\gamma(X) = 0) \neq \mathbb{P}_1(\gamma(X) = 0)$;
2) every node at level 1 uses a LLRQ, with a threshold t that satisfies $e_{0,\gamma} < t < e_{1,\gamma}$;
3) all other nodes other than the fusion center, use the majority rule: send a 1 if and only if more than half of the received messages are equal to 1;
4) the fusion center uses a LLRQ with threshold $t = -D(\eta) + \epsilon$.

Proof: (Outline): Similar to the proof of Proposition 1, we first lower bound the optimal error exponent. Consider the fusion center f. Suppose a genie tells each $v \in C(f)$ the true hypothesis and each node v sends this information to the fusion center. Because of the BSC from each node v to f, the received message at f has distribution $\mathrm{Bern}(\eta)$ under H_0, and $\mathrm{Bern}(1-\eta)$ under H_1. From Stein's Lemma [25], the optimal error exponent is $-D(\eta)$. The performance in the absence of the genie cannot be better. Therefore

$$\liminf_{n\to\infty} \frac{1}{|C(f)|} \log\beta^* \ge -D(\eta). \tag{8}$$

We now turn to the proof of the upper bound. Consider the strategy described in the proposition. Let v be a node at level 1. This node v receives a message Z_u from each leaf $u \in C(v)$. These messages are binary, conditionally i.i.d., but with a different distribution under each hypothesis. Moreover, v receives at least c_n such messages. In such a case, it is well known [25] (and also easy to show from laws of large numbers) that if the node v uses a LLRQ with a threshold t that satisfies $e_{0,\gamma} < t < e_{1,\gamma}$, then the error probabilities at node v decay exponentially fast with c_n; that is, there exist some Δ and $\delta > 0$ such that

$$\mathbb{P}_0(Y_v = 1) \le \Delta e^{-c_n\delta}, \quad \mathbb{P}_1(Y_v = 0) \le \Delta e^{-c_n\delta} \quad \forall n. \tag{9}$$

Taking into account the statistics of the BSC, we have

$$\mathbb{P}_0(Z_v = 1) \le \eta + \Delta e^{-c_n\delta},\ \mathbb{P}_1(Z_v = 0) \le \eta + \Delta e^{-c_n\delta}\ \forall n. \tag{10}$$

In particular, for n sufficiently large, and for all level 1 nodes v, we have $\mathbb{P}_0(Z_v = 1) < 1/2$ and $\mathbb{P}_1(Z_v = 0) < 1/2$. Consider now a node w at level 2. This node receives at least c_n independent messages Z_v from each $v \in C(w)$, where these messages have error probabilities $\mathbb{P}_0(Z_v = 1) < 1/2$ and $\mathbb{P}_1(Z_v = 0) < 1/2$. The node w then uses a majority rule to form its message Y_w. It is easy to show, using laws of large numbers, that (9) holds for Y_w, with possibly different constants Δ and δ. Then, (10) also holds for Z_w. Continuing inductively, we conclude that there exist constants $\Delta > 0$ and $\delta > 0$, such that for all nodes v, (10) holds.

Consider now the fusion center, and a typical node $v \in C(f)$. From (10), if n is sufficiently large, the message Z_v received by the fusion center has KL divergence at least $D(\eta) - \epsilon/2$ [note that $D(\cdot)$ is continuous and decreasing over (0,1/2)]. It then follows, from Cramér's Theorem [25], that the Type II error exponent at the fusion center is less than or equal to $-D(\eta) + \epsilon$, if a LLRQ with threshold $t = -D(\eta) + \epsilon$ is used at the fusion center. Moreover, the Type I error exponent is strictly negative in this case, so that the Type I error probability can be brought to below α when n is sufficiently large. The proof is now complete. ■

C. Discussion

We have established that the detection performance of a tree network in which the communication channel between two nodes is a BSC, and which has a height $h \geq 2$, is the same as if every immediate predecessor of the fusion center had perfect knowledge of the true hypothesis. On the other hand, when compared to the case of reliable communications (where the error probability falls exponentially with the number of nodes [15]), the performance is significantly degraded. Thus, channel noise can be detrimental.

Consider a tree network in which all nonleaf nodes have the same number of immediate predecessors c_n. Suppose that each node estimates its channel to its immediate successor, and sends its message only if that message will be received reliably. In this case, the number of immediate predecessors of a node of level $k \geq 1$ has a Binomial distribution $\mathcal{B}(c_n, 1-\eta)$ with mean $\lambda_k = c_n(1-\eta)$. In Section III, we showed that the Type II error probability, when the network is operating in this manner, falls exponentially with $\lambda(h) = c_n^h(1-\eta)^h$. On the other hand, Proposition 3 shows that the minimum error probability achievable when messages are sent regardless of channel conditions, falls exponentially with c_n. Hence, our results suggest that in a dense sensor network of height $h \geq 2$, if a node determines that it cannot reliably transmit its message to its immediate successor, it is better for the node to remain silent. Our results also suggest that when designing a large scale sensor network, it is more important to ensure that there is reliable communication between nodes (e.g., by using sufficient transmission power), than to guard against node failures.

D. Error Exponent With Small Channel Crossover Probabilities

In Proposition 3, we showed that the Type II error probability decays exponentially fast with $|C(f)|$, when the channel error probability η is fixed. In this section, we let η go to zero as n increases, which corresponds to increasing the transmit power of each node.[6] Under an assumption on the rate at which η goes to zero, we show that the Type II error probability can be made to decay exponentially fast with n, at rate g_P^*.

Proposition 4: Suppose that Assumptions 1 and 2 hold. Suppose also the following.

1) If $h=1$, then $\lim_{n\to\infty} \eta = 0$.
2) Let $l_M = \max_{v\in C(f)} l(v)$. If $h \geq 2$, then $\limsup_{n\to\infty}(1/l_M)\log\eta \leq g_P^*$.

Fix an $\epsilon \in (0, -g_p^*/h)$. Suppose that all leaves use the same transmission function $\gamma \in \Gamma$, chosen so that $-\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) < g_P^* + \epsilon$. Suppose also that each level k $(k \geq 1)$ node v sends a message 0 iff $S_v/l(v) \leq t_k := g_P^* + k\epsilon$. Then, for n sufficiently large, we have for every level k node v

$$\frac{1}{l(v)} \log \mathbb{P}_1(Y_v = 0) \leq t_k \tag{11}$$

$$\frac{1}{l(v)} \log \mathbb{P}_0(Y_v = 1) \leq -\epsilon_k < 0 \tag{12}$$

where $\epsilon_1, \ldots, \epsilon_h$ are positive reals less than or equal to ϵ. In particular, for any $h \geq 1$ and $\alpha \in (0,1)$, the optimal error exponent is

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^* = g_P^*.$$

Proof: If $h=1$, the situation is similar to the case considered in Section IV-A. As $\eta \to 0$, $e_{0,\gamma}$ approaches $-\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma)$, and e_0 approaches g_P^*, which leads to the desired result. The details are omitted.

We now consider the case where $h \geq 2$. From the Chernoff bound, we have

$$\frac{1}{l(v)} \log \mathbb{P}_1\left(\frac{S_v}{l(v)} \leq t_k\right) \leq \frac{1}{l(v)} \log\left(e^{l(v)t_k}\mathbb{E}_1[e^{-S_v}]\right) = t_k$$

hence (11) follows. To show the inequality (12), we proceed by induction on k. When $k=1$, the inequality follows from Cramér's Theorem [25]. Suppose that (12) holds for all level k nodes. Consider a level $k+1$ node v. For any $s \in [0,1]$, we have from the Chernoff bound

$$\begin{aligned}
&\frac{1}{l(v)} \log \mathbb{P}_0\left(\frac{S_v}{l(v)} > t_{k+1}\right)\\
&\leq -st_{k+1} + \frac{1}{l(v)} \log \mathbb{E}_0\left[\exp(sS_v)\right]\\
&= -st_{k+1} + \frac{1}{l(v)} \sum_{u\in C(v)} \log \mathbb{E}_0\left[\exp(s\bar{L}_u)\right]\\
&\leq -st_{k+1} + \frac{1}{l(v)}\\
&\quad \times \sum_{u\in C(v)} \log\left\{\mathbb{P}_1(Z_u=0)^s + \mathbb{P}_0(Z_u=1)^{1-s}\right\}\\
&\leq -st_{k+1} + \frac{1}{l(v)}\\
&\quad \times \sum_{u\in C(v)} \max\left\{s\log\mathbb{P}_1(Z_u=0),\right.\\
&\qquad\qquad \left.(1-s)\log\mathbb{P}_0(Z_u=1)\right\} + \frac{|C(v)|}{l(v)}\log 2\\
&\leq -st_{k+1} + \max\left\{st_k, -(1-s)\epsilon_k\right\} + \frac{2|C(v)|}{l(v)}\log 2. \qquad (13)
\end{aligned}$$

The last inequality follows because

$$\begin{aligned}
s\log\mathbb{P}_1(Z_u=0) &\leq s\log\left(\mathbb{P}_1(Y_u=0)+\eta\right)\\
&\leq s\log\left(e^{l(u)t_k} + e^{l(u)\left(g_P^*+\epsilon\right)}\right)\\
&\leq l(u)st_k + \log 2.
\end{aligned}$$

(In the penultimate inequality, we used (11), and the assumption on the decay rate of η; in the last inequality, we used the fact $g_P^* + \epsilon \leq t_k$.) Similarly, using the induction hypothesis instead of (11), we have

$$(1-s)\log\mathbb{P}_0(Z_u=1) \leq -l(u)(1-s)\epsilon_k + \log 2$$

hence inequality (13) holds. We choose s in the right-hand side of (13) so that $st_k = -(1-s)\epsilon_k$. Note that $t_k < 0$ and $\epsilon_k > 0$, which together guarantee that $0 < s < 1$. Recall that every

[6] We suppress the dependence of η on n in the notation.

nonleaf node is assumed to have degree at least c_n, which grows to infinity. Thus, for n sufficiently large, (13) implies that

$$\begin{aligned}\frac{1}{l(v)}\log \mathbb{P}_0(Y_v=1) &\leq -s\epsilon + \frac{2\,|C(v)|}{l(v)}\log 2\\ &\leq -s\epsilon + \frac{2}{c_n}\log 2\\ &\leq -s\epsilon/2 := -\epsilon_{k+1}\end{aligned}$$

hence (12) holds for level $k+1$ nodes. The induction is now complete.

To complete the proof of the proposition, since $l(f)/n \to 1$ as $n\to\infty$ (cf. Lemma 4 of [15]), (11) yields

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta^* = \limsup_{n\to\infty}\frac{1}{l(f)}\log\beta^* \leq g_p^* + h\epsilon$$

while (12) ensures that the Type I error probability is less than α for n sufficiently large. Finally, the optimal error exponent is obtained by letting ϵ go to 0, and the proposition is proved. ■

Using a similar argument as in the proof of the previous proposition, it can be shown that the condition

$$\limsup_{n\to\infty}\frac{1}{l_M}\log\eta < 0 \tag{14}$$

is sufficient for a tree network of height $h\geq 2$ to achieve a Type II error probability that decays exponentially fast with n, although the error exponent can be worse (less negative) than g_P^*.

E. Energy Efficiency Comparison

In this subsection, we consider $l(f)$ nodes arranged on a grid, with neighboring nodes unit distance apart. The $l(f)$ nodes are the leaves of our network, but we are otherwise free to configure the network, and to possibly introduce additional nodes that will serve as message relays. We will compare the energy consumption of a parallel configuration with that of a tree network of height $h\geq 2$, under the assumptions of Proposition 4. In both cases, the fusion center is placed at the center of the entire grid.

To construct a tree of height h, we add new nodes at levels $1,\ldots,h-1$, as follows (see Fig. 1). Let r be a positive integer which is a perfect square. Partition the grid of nodes into r equal sub-squares, each of which is called a level $h-1$ sub-square. At the center of each sub-square, we place a new node, which serves as a level $h-1$ node. Next, partition each level $h-1$ sub-square into r further sub-squares, and place a new node at the center of each of the latter sub-squares. These are the level $h-2$ nodes, which send their messages to the level $h-1$ node of that sub-square. We repeat this procedure $h-1$ times. Finally, all the leaves in a level 1 sub-square send their messages to the level 1 node in that sub-square.

The total number of nodes is $n = l(f) + (r^h-1)/(r-1)$. As we consider progressively larger values of $l(f)$, we adjust the value of r used in the previous construction, so that $l(f)/r^{h-1}\to\infty$ and $l(f)/r^h \to 0$, as $n\to\infty$. We compare the performance and energy consumption of this tree network with that of a parallel configuration in which all $l(f)$ nodes send their messages directly to the fusion center. (Since $l(f)/n\to 1$ as $n\to\infty$, the results would also be the same for a parallel configuration with n, instead of $l(f)$, nodes.)

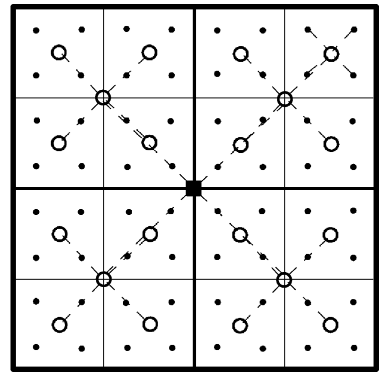

Fig. 1. Tree network of height 3, with $r=4$. The circles represent the new nodes that we have added. The dotted lines indicate communication links. Only one level 1 sub-square (the top right one) is shown with all its communication links.

In the tree network ($h\geq 2$) that we have constructed, the condition

$$\limsup_{n\to\infty}\frac{1}{l_M}\log\eta \leq g_P^*$$

is not only sufficient, but also necessary for the Type II error exponent to be g_P^*. To see this, suppose that $Z_1,\ldots,Z_r$ are messages received at the fusion center. For the Type I error constraint to be satisfied, there exists $(z_1,\ldots,z_r)$ such that $\gamma_f(z_1,\ldots,z_r)=0$. Moreover, for any $Y_1,\ldots,Y_r\in\{0,1\}$, we have $\mathbb{P}_1(Z_1=z_1,\ldots,Z_r=z_r|Y_1,\ldots,Y_r)\geq\eta^r$. Therefore, we obtain

$$\begin{aligned}\beta^* &\geq \mathbb{P}_1(Z_1=z_1,\ldots,Z_r=z_r)\\ &= \mathbb{E}_1\left[\mathbb{P}_1(Z_1=z_1,\ldots,Z_r=z_r|Y_1,\ldots,Y_r)\right]\geq\eta^r.\end{aligned}$$

Hence, if $\limsup_{n\to\infty}(1/l_M)\log\eta > g_P^*$, we would have $\lim_{n\to\infty}(1/l(f))\log\beta^* > g_P^*$, since $l_M = l(f)/r$.

We assume that each node employs antipodal signalling, and the received signal is corrupted by additive white Gaussian noise with variance $N_0/2$: a node receives a $N(\sqrt{E_b}, N_0/2)$ random variable if a 1 is sent by its immediate predecessor and a $N(-\sqrt{E_b}, N_0/2)$ random variable if a 0 is sent. The recipient node performs a maximum a posterior probability test to determine if a 1 or 0 was sent. The resulting channel error probability is

$$\eta = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\approx\frac{1}{2}\sqrt{\frac{N_0}{E_b\pi}}e^{-E_b/N_0}$$

where $Q(\cdot)$ is the Gaussian complementary error function. To satisfy the conditions in Proposition 4, we choose E_b as follows:

1) if $h=1$, let $E_b=E(n)$, where $E(n)\to\infty$ as $n\to\infty$;
2) if $h\geq 2$, let $E_b = c\cdot l(f)/r$, where $c\geq -g_P^*/N_0$ is a constant.

We also assume a path-loss model, so that the received bit energy at each receiver node is $E_b = E_0/D^a$, where D is the transmission distance, a is the path-loss exponent, and E_0 is the transmission energy expended by the transmitting node. Therefore, the transmission energy of a node is $E_0 = E_bD^a$. In line with standard wireless communications models [26], we take $2\leq a\leq 4$.

Let E_{CT} be the circuit processing energy required by each node, and E_{CR} be the receiver circuit energy incurred by a receiver node per message received [27]. The total energy E_P expended by a parallel configuration is given below. The first term is the receiver circuit energy of the fusion center, the second term is the processing energy expended by all the nodes, the third term is the total transmission energy, and $c(n)$ is the average path-loss D^a suffered by the nodes. We have

$$E_P = l(f)E_{CR} + (l(f)+1)\,E_{CT} + l(f)E(n)c(n).$$

Since more than half of the nodes are at distance at least $\sqrt{l(f)}/4$ from the fusion center, we obtain[7]

$$\begin{aligned} E_P &\geq l(f)(E_{CR}+E_{CT}) + l(f)E(n)\cdot\frac{1}{2}\left(\frac{\sqrt{l(f)}}{4}\right)^a \\ &= \Omega\left(n^{1+a/2}E(n)\right). \end{aligned}$$

For the tree network with height $h \geq 2$, we have the following upper bound on the total energy consumption E_T. The first term is the total processing energy of all the nodes, the second term is the receiver circuit energy expended by nodes from level 1 to level h, the third term is an upper bound on the transmission energy expended by nodes from level 1 to level $h-1$, and the last term is an upper bound on the transmission energy expended by the leaves. So, we have

$$\begin{aligned} E_T &\leq nE_{CT} + (n-1)E_{CR} + E_b\sum_{k=1}^{h-1} r^k\left(\frac{\sqrt{l(f)}}{r^{\frac{k-1}{2}}}\right)^a \\ &\quad + l(f)E_b\left(\frac{\sqrt{l(f)}}{r^{\frac{h-1}{2}}}\right)^a \\ &\leq n(E_{CT}+E_{CR}) + c\frac{l(f)}{r}\sum_{k=1}^{h-1} r^k\cdot\frac{n^{a/2}}{r^{k-1}} \\ &\quad + l(f)\cdot c\frac{l(f)}{r}\cdot\frac{n^{a/2}}{r^{h-1}} \\ &\leq n(E_{CT}+E_{CR}) + c(h-1)n^{1+a/2} + cn^{1+a/2}\frac{l(f)}{r^h} \\ &= O(n^{1+a/2}). \end{aligned}$$

The above analysis shows that for large n, $E_T < E_P$. Hence, the tree network consumes less energy than the parallel configuration, if both networks are designed to have the same error exponent g_P^*.

V. The Bayesian Problem

In this section, we consider the Bayesian formulation of the problems analyzed in Sections III and IV, under some additional simplifying assumptions.

Suppose that we are given positive prior probabilities π_0 and π_1 for each hypothesis. Let $P_e = \pi_0\mathbb{P}_0(Y_f = 1) + \pi_1\mathbb{P}_1(Y_f = 0)$ be the probability of error at the fusion center, and let P_e^* be the minimum probability of error, where the minimization is over all strategies. We assume that the fusion center always uses the optimal fusion rule, namely the maximum a posteriori probability rule. In this section, we assume that all nodes are constrained to sending 1-bit messages. We also make the following assumption on the observations at the leaves.

Assumption 4: Either one of the following holds.

i) The observations X_i at the leaves take values in a finite set.

ii) Assumption 2 and the condition $\mathbb{E}_1[\log^2(\mathrm{d}\mathbb{P}_1^X/\mathrm{d}\mathbb{P}_0^X)] < \infty$ hold.

For each $\gamma \in \Gamma$, let

$$\Lambda(\gamma) = \min_{s\in[0,1]} \log\mathbb{E}_0\left[\left(\frac{\mathrm{d}\mathbb{P}_1^\gamma}{\mathrm{d}\mathbb{P}_0^\gamma}\right)^s\right].$$

Under Assumptions 1 and 4, it is known that the optimal error exponent for a parallel configuration with a deterministic number of nodes is given by

$$\Lambda^* = \inf_{\gamma\in\Gamma}\Lambda(\gamma).$$

According to Propositions 2 and 3 of [19], Assumptions 1 and 4 imply the following lemma.

Lemma 5: Suppose that Assumptions 1 and 4 hold. Then, for any choice of transmission functions $\gamma_1,\ldots,\gamma_n$ used by the n leaves in a parallel configuration, the resulting probability of error, $P_e(n)$, assuming that all transmissions are reliable, satisfies

$$P_e(n) = \exp\left\{\sum_{i=1}^n \Lambda(\gamma_i) + f(n)\right\}$$

where $f(n)$ is a function such that $\lim_{n\to\infty} f(n)/n = 0$.

In the next two subsections, we consider separately the cases of node failures and unreliable communications, in the Bayesian framework.

A. Node Failures

For tractability, we consider only the case where for all $k \geq 1$, μ_k is the Poisson distribution with mean λ_k. We have the following proposition, which yields the optimal error exponent in the presence of node failures. Unlike the Neyman–Pearson case, where the Type II error probability decays exponentially fast with the expected number, $\lambda(h)$ of leaves, the Bayesian error probability decays exponentially with λ_h.

Proposition 5: Suppose that Assumptions 1 and 4 hold.

a) If $h = 1$, the optimal error exponent is given by

$$\lim_{\lambda_1\to\infty}\frac{1}{\lambda_1}\log P_e^* = -1 + e^{-\Lambda^*}.$$

b) If $h \geq 2$, the optimal error exponent is given by

$$\lim_{\lambda^*\to\infty}\frac{1}{\lambda_h}\log P_e^* = -1.$$

Furthermore, the optimal error exponent remains unchanged if we restrict all leaves to use the same transmission function $\gamma \in \Gamma$, and all other nodes to use a majority rule.

Proof: 1) Suppose that $h = 1$. For every n, we have from Lemma 5, $P_e(n) \geq \exp\{n\Lambda^* + f(n)\}$. Furthermore, $P_e = \mathbb{E}[P_e(N)]$, where N has a Poisson distribution with mean λ_1. Fix some $\epsilon > 0$. Let n_0 be such that $|f(n)| \leq n\epsilon$, for every

[7] For two nonnegative functions f and g, we write $f(n) = \Omega(g(n))$ (respectively, $f(n) = O(g(n))$) if for all n sufficiently large, there exists a positive constant c such that $f(n) \geq cg(n)$ (resp. $f(n) \leq cg(n)$).

$n > n_0$. Let $m = \sup_{1 \le n \le n_0}\{|f(n)|\}$, and notice that $|f(n)| \le m + n\epsilon$. We have

$$P_e \ge \sum_{n=0}^{\infty} e^{-\lambda_1}\frac{\lambda_1^n}{n!}e^{n(\Lambda^*-\epsilon)-m} = e^{-m-\lambda_1}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\lambda_1 e^{\Lambda^*-\epsilon}\right)^n = \exp\left\{\lambda_1(e^{\Lambda^*-\epsilon}-1)-m\right\}.$$

Therefore,

$$\liminf_{\lambda_1\to\infty}\frac{1}{\lambda_1}\log P_e^* \ge -1 + e^{\Lambda^*-\epsilon}.$$

Since ϵ was arbitrary, it follows that

$$\liminf_{\lambda_1\to\infty}\frac{1}{\lambda_1}\log P_e^* \ge -1 + e^{\Lambda^*}.$$

For a corresponding upper bound, let all leaves use a transmission function γ^* such that $\Lambda(\gamma^*) \le \Lambda^* + \epsilon$. We then have

$$P_e^* \le \sum_{n=0}^{\infty} e^{-\lambda_1}\frac{\lambda_1^n}{n!}e^{n(\Lambda^*+\epsilon)+f(n)} \le \sum_{n=0}^{\infty} e^{-\lambda_1}\frac{\lambda_1^n}{n!}e^{n(\Lambda^*+2\epsilon)+m} = \exp\{\lambda_1(e^{\Lambda^*+2\epsilon}-1)+m\}.$$

We take logarithms, divide by λ_1, and take the limit as $\lambda_1 \to \infty$. Using also the fact that ϵ was arbitrary, we obtain that

$$\limsup_{\lambda_1\to\infty}\frac{1}{\lambda_1}\log P_e^* \le -1 + e^{\Lambda^*}.$$

2) (Outline) Suppose now that $h \ge 2$. Notice that there is a probability $e^{-\lambda_h}$ that the fusion center has no predecessors, and a further probability of $\min\{\pi_0, \pi_1\}$ of making an error, so that $P_e^* \ge \min\{\pi_0,\pi_1\}e^{-\lambda_h}$. It follows that $\limsup_{\lambda^*\to\infty}(1/\lambda_h)\log P_e^* \ge -1$.

For a corresponding upper bound, consider the case where all leaves use the same transmission function, and all other nodes use a majority rule. An easy induction argument shows that for every immediate predecessor of the fusion center, $\mathbb{P}_0(Y_u = 1)$ and $\mathbb{P}_1(Y_u = 0)$ can be brought arbitrarily close to zero, as $\lambda^* \to \infty$. This brings us to a situation similar to the one considered in part (a), except that now Λ^* can be replaced by an arbitrarily negative constant. A calculation similar to the one in part (a) yields $\limsup_{\lambda_h\to\infty}(1/\lambda_1)\log P_e^* \le -1$. ■

B. Unreliable Communications

In the case of unreliable communications, the corresponding results are obtained easily.

Proposition 6: Suppose that Assumptions 1 and 4 hold.

i) If $h = 1$, it is optimal to have all leaves use the same transmission function, and the optimal error exponent is given by

$$\lim_{n\to\infty}\frac{1}{|C(f)|}\log P_e^* = \inf_{\gamma\in\Gamma}\min_{s\in[0,1]}\log\left(\sum_{z=0}^{1}\left(q_0^{\gamma}(z)\right)^{1-s}\left(q_1^{\gamma}(z)\right)^{s}\right).$$

ii) For $h \ge 2$, it is optimal to have all leaves use the same transmission function γ, where γ is chosen so that $\Lambda(\gamma) < 0$, and to have all intermediate nodes use a majority rule. Furthermore, the optimal error exponent is given by

$$\lim_{n\to\infty}\frac{1}{|C(f)|}\log P_e^* = \frac{1}{2}\log\left(4\eta(1-\eta)\right).$$

Proof: (Outline): Part (i) follows from Theorem 1 of [19]. As for part (ii), an argument similar to the proof of Proposition 3 shows that the probability of error at each intermediate node converges to zero, so that the messages received by the fusion center have asymptotic distributions $\mathrm{Bern}(\eta)$ or $\mathrm{Bern}(1-\eta)$, under H_0 or H_1, respectively. The final result then follows immediately from Chernoff's bound [25]. ■

VI. Conclusion

We have studied the effects of node failures and unreliable communications in a dense sensor network, arranged as a tree of bounded height. We have analyzed the asymptotically optimal performance in order to gain insights into otherwise intractable problems. Our analysis suggests that, in practice, it is preferable to have a node faced with an unreliable channel remain silent (as if it had failed). It also suggests that, when designing a large scale sensor network, it is more important to ensure that nodes can communicate reliably with each other (e.g., by boosting the transmission power) than to ensure that nodes are robust to failures.

We now discuss some future research directions. Our assumption that the leaves make (conditionally) i.i.d. observations is restrictive and will often be violated. The case of correlated observations in tree networks has been unexplored. For some recent work in the case of a parallel configuration, we refer the reader to [28] and [29].

This paper concentrated on the case of trees with fixed or bounded height. It would be of interest to understand the dependence of the error probability on the height of the network. For the case of failure-proof sensors and reliable transmissions, we have shown that the optimal Bayesian error probability in a tandem configuration decays sub-exponentially [30]. A similar problem is worth studying for the case of general tree networks with node failures and/or unreliable communications.

Other than node failures and unreliable communications, another threat to a sensor network is malicious tampering of some nodes so that they report false information to the fusion center [31], [32]. It would be of interest to characterize the impact of such Byzantine sensors on the detection performance.

APPENDIX

A. Proof of Lemma 1

From Markov's Inequality

$$\mathbb{P}(X > a/c_1) \leq c_1, \quad \mathbb{P}(Y > b/c_2) \leq c_2.$$

Therefore, by the union bound, we have

$$\mathbb{P}\left((\{X > a/c_1\} \cup \{Y > b/c_2\}) \cap A\right) \leq c_1 + c_2.$$

This implies that

$$\mathbb{P}\left(\{X \leq a/c_1\} \cap \{Y \leq b/c_2\} \cap A\right) \geq \mathbb{P}(A) - c_1 - c_2 > 0.$$

Hence, there exists some $\omega \in A$ such that $X(\omega) \leq a/c_1$ and $Y(\omega) \leq b/c_2$.

B. Proof of Lemma 2

a) We use induction on h. For $h = 1$, (a) follows from Assumption 3. Suppose that the claim holds for GW trees of height $h-1$, and consider a GW tree of height h. Recall that N_f is the cardinality of the set $C(f)$ of immediate predecessors of the fusion center f. For $u \in C(f)$, we observe that $l(u)$ is the number nodes in a GW tree of height $h-1$, rooted at u. The induction hypothesis yields $\mathbb{E}[l(u)] = \lambda(h-1)$ and $\mathrm{var}[l(u)]/\lambda^2(h-1) \to 0$, as $\lambda^* \to \infty$. Furthermore, the random variables $l(u)$ are i.i.d. Let w be a typical element of $C(f)$. We have

$$\mathbb{E}\left[l(f)\right] = \mathbb{E}\left[\sum_{u \in C(f)} l(u)\right] = \mathbb{E}[N_f]\mathbb{E}\left[l(w)\right] = \lambda(h).$$

Using a well-known formula for the variance of the sum of a random number of i.i.d. random variables, we also have

$$\begin{aligned}\frac{\mathrm{var}\left[l(f)\right]}{\lambda^2(h)} &= \frac{\mathrm{var}[N_f]\left(\mathbb{E}\left[l(w)\right]\right)^2 + \mathbb{E}[N_f]\mathrm{var}\left[l(w)\right]}{\lambda^2(h)} \\ &= \frac{\mathrm{var}[N_f]}{\lambda_h^2} \cdot \frac{\lambda^2(h-1)}{\lambda^2(h-1)} + \frac{\lambda_h}{\lambda_h^2} \cdot \frac{\mathrm{var}\left[l(w)\right]}{\lambda^2(h-1)}\end{aligned}$$

which, using the induction hypothesis and Assumption 3, converges to 0.

b) This is an immediate consequence of Chebychev's Inequality.

ACKNOWLEDGMENT

The first author wishes to acknowledge helpful discussions with Tony Quek. The authors would also like to thank the anonymous reviewers for their careful reading of this manuscript.

REFERENCES

[1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

[2] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.

[3] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 282–292, Mar. 1990.

[4] J. N. Tsitsiklis, "Decentralized detection," *Adv. Stat. Signal Process.*, vol. 2, pp. 297–344, 1993.

[5] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[6] B. Chen and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

[7] L. K. Ekchian and R. R. Tenney, "Detection networks," in *Proc. IEEE Conf. Dec. Control*, 1982, pp. 686–691.

[8] A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 6, pp. 789–797, Dec. 1987.

[9] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part I—Tandem structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sep./Oct. 1991.

[10] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part II—Tree structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 1, pp. 211–221, Jan./Feb. 1993.

[11] J. D. Papastavrou and M. Athans, "On optimal distributed decision architectures in a hypothesis testing environment," *IEEE Trans. Autom. Control*, vol. 37, no. 8, pp. 1154–1169, Aug. 1992.

[12] A. Pete, K. Pattipati, and D. Kleinman, "Optimization of detection networks with multiple event structures," *IEEE Trans. Autom. Control*, vol. 39, no. 8, pp. 1702–1707, Aug. 1994.

[13] S. Alhakeem and P. K. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 9–20, Jan. 1995.

[14] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," in *Proc. Data Compression Conf.*, Mar. 2007, pp. 243–252.

[15] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?," *IEEE Trans. Inf. Theory*, 2007, accepted for publication.

[16] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 4, pp. 462–472, 2005.

[17] T. Q. S. Quek, D. Dardari, and M. Z. Win, "Energy efficiency of dense wireless sensor networks: To cooperate or not to cooperate," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 459–469, 2007.

[18] A. Anandkumar and L. Tong, "A large deviation analysis of detection over multi-access channels with random number of sensors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. 4, pp. 1097–1100.

[19] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, pp. 167–182, 1988.

[20] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[21] A. Kashyap, "Comments on "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels"," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1274–1275, Mar. 2006.

[22] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3349–3358, Jul. 2006.

[23] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.

[24] R. Durrett, *Probability: Theory and Examples*, 2nd ed. New York: Duxbury Press, 1995.

[25] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer-Verlag, 1998.

[26] D. N. C. Tse and P. Viswanath, *Fundatmentals of Wireless Communications*. New York: Cambridge University Press, 2005.

[27] E. Björnemo, M. Johansson, and A. Ahlén, "Two hops is one too many in an energy-limited wireless sensor network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2007, pp. 181–184.

[28] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[29] W. Li and H. Dai, "Distributed detection in large-scale sensor networks with correlated sensor observations," in *Proc. Allerton Conf. Commun., Control, Comput.*, Sep. 2005.

[30] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection probabilities in long tandems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2007, pp. 837–840.

[31] K. Kursawe, "Optimistic Byzantine agreement," in *Proc. IEEE Symp. Reliable Distrib. Syst.*, Oct. 2002, pp. 262–267.

[32] S. Marano, V. Matta, and L. Tong, "Distributed inference in the presence of Byzantine sensors," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2006, pp. 281–284.

Wee Peng Tay received the B.S. degree in electrical engineering and mathematics, the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2002, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2008.

His research interests include distributed decision making in sensor networks, data fusion, distributed algorithms, stochastic networks, and applied probability.

John N. Tsitsiklis (F'99) received the B.S. degree in mathematics and the B.S., M.S., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1980, 1980, 1981, and 1984, respectively.

He is currently a Clarence J. Lebel Professor with the Department of Electrical Engineering, MIT. He has also served as a Codirector of the MIT Operations Research Center from 2002 to 2005. His research interests include the fields of systems, optimization, communications, control, and operations research. He has coauthored four books and more than a hundred journal papers in these areas.

Prof. Tsitsiklis was a recipient of an Outstanding Paper Award from the IEEE Control Systems Society (1986), the M.I.T. Edgerton Faculty Achievement Award (1989), the Bodossakis Foundation Prize (1995), and the INFORMS/CSTS Prize (1997). He is currently a member of the editorial board for the Springer-Verlag "Lecture Notes in Control and Information Sciences" series, an Associate Editor of *Mathematics of Operations Research*, a member of the National Academy of Engineering, and a member of the National Council on Research and Technology in Greece.

Moe Z. Win (S'85–M'87–SM'97–F'04) received the B.S. degree (*magna cum laude*) from Texas A&M University, College Station, in 1987, and the M.S. degree from the University of Southern California (USC), Los Angeles, in 1989, both in electrical engineering. As a Presidential Fellow at USC, he received both the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering, in 1998.

He is an Associate Professor with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology (MIT), Cambridge. Prior to joining MIT, he spent five years with AT&T Research Laboratories and seven years with the Jet Propulsion Laboratory. His main research interests are the applications of mathematical and statistical theories to communication, detection, and estimation problems. Specific current research topics include measurement and modeling of time-varying channels, design and analysis of multiple antenna systems, ultra-wide bandwidth (UWB) systems, optical transmission systems, and space communications systems.

Prof. Win has been actively involved in organizing and chairing a number of international conferences. He served as the Technical Program Chair for the IEEE Conference on Ultra Wideband in 2006, the IEEE Communication Theory Symposia of ICC-2004 and Globecom-2000, and the IEEE Conference on Ultra Wideband Systems and Technologies in 2002; Technical Program Vice-Chair for the IEEE International Conference on Communications in 2002; and the Tutorial Chair for the IEEE Semiannual International Vehicular Technology Conference in Fall 2001. He served as the chair (2004-2006) and secretary (2002-2004) for the Radio Communications Committee of the IEEE Communications Society. He is currently an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He served as Area Editor for *Modulation and Signal Design* (2003-2006), Editor for *Wideband Wireless and Diversity* (2003-2006), and Editor for *Equalization and Diversity* (1998-2003), all for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was Guest-Editor for the 2002 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on Ultra -Wideband Radio in Multiaccess Wireless Communications).

Prof. Win received the International Telecommunications Innovation Award from Korea Electronics Technology Institute in 2002, a Young Investigator Award from the Office of Naval Research in 2003, and the IEEE Antennas and Propagation Society Sergei A. Schelkunoff Transactions Prize Paper Award in 2003. In 2004, Dr. Win was named Young Aerospace Engineer of the Year by AIAA, and garnered the Fulbright Foundation Senior Scholar Lecturing and Research Fellowship, the Institute of Advanced Study Natural Sciences and Technology Fellowship, the Outstanding International Collaboration Award from the Industrial Technology Research Institute of Taiwan, and the Presidential Early Career Award for Scientists and Engineers from the United States White House. He was honored with the 2006 IEEE Eric E. Sumner Award "for pioneering contributions to ultra-wide band communications science and technology." Professor Win is an IEEE Distinguished Lecturer and elected Fellow of the IEEE, cited "for contributions to wideband wireless transmission."