Bayesian Detection in Bounded Height Tree Networks

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

***Abstract*—We study the detection performance of large scale sensor networks, configured as trees with bounded height, in which information is progressively compressed as it moves towards the root of the tree. We show that, under a Bayesian formulation, the error probability decays exponentially fast, and we provide bounds for the error exponent. We then focus on the case where the tree has certain symmetry properties. We derive the form of the optimal exponent within a restricted class of easily implementable strategies, as well as optimal strategies within that class. We also find conditions under which (suitably defined) majority rules are optimal. Finally, we provide evidence that in designing a network it is preferable to keep the branching factor small for nodes other than the neighbors of the leaves.**

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

I. Introduction

We consider a sensor network, configured as a directed tree, with a fusion center at its root. The objective of the network is to make a decision between two given hypotheses H_0 and H_1. Observations are obtained at the nodes of the tree, and information is propagated from the leaves towards the root. However, because of resource constraints, e.g., a restriction to single-bit messages, every node is required to compress or quantize its information (its observation and the messages it has received) before forming its own message. Based on the received information, the root or fusion center makes a decision about the true hypothesis. Our objective is to understand the scaling of the error probability at the fusion center, as the number of nodes increases, and its dependence on qualitative properties of the tree.

In the well-studied parallel configuration (see e.g., [1]–[10]), each node sends its compressed information directly to the fusion center. A tree, on the other hand, allows for shorter-range communications, thus making better use of communication resources. Tree networks have been studied in several references, such as [11]–[18]. It is known that under the assumptions to be made in this paper (conditioned on either hypothesis, the observations at the different nodes are i.i.d.), optimal quantization strategies take the form of likelihood-ratio quantizers, and one can obtain "person-by-person optimality" conditions that need to be satisfied by the optimal quantizers. Nevertheless, finding the optimal quantizers, and hence characterizing the detection performance, is a rather intractable problem even for a moderate number of nodes. For this reason, in the spirit of [19], we focus on the exponential rate of decay of error probabilities.

In [20], we studied the Neyman–Pearson variant of the problem considered in this paper. We showed that the error probability decays exponentially fast with the number of nodes (this is apparently not the case when the height is unbounded, e.g., in a tandem configuration [21]–[24]); furthermore, in some cases the error exponent associated with a tree configuration turned out to be the same as for the parallel configuration. In this paper, we continue this investigation by focusing on the Bayesian formulation. Similar to the Neyman–Pearson case, we will see that for bounded height trees error probabilities decay exponentially fast. However, the optimal error exponent is generically worse than the one associated with a parallel configuration (cf. Proposition 2), and is also harder to characterize exactly. In order to make further progress, we place some additional restrictions on the trees to be considered, as well as on the allowed quantization strategies. The following example serves to motivate some of our assumptions.

Example 1: (Random Nodes in the Unit Square) Suppose that we distribute n nodes randomly in the unit square and place a fusion center at the center of the square. We are interested in configuring the nodes so that every node is at most two hops away from the fusion center.

One possibility (to be referred to as Design I) is to fix some m, and divide the square into m sub-squares, each with side of length $1/\sqrt{m}$ (see Fig. 1). For large n, there are approximately n/m nodes in each of these sub-squares. We let all nodes within a sub-square transmit their messages to an "aggregator" node in that sub-square. In this way, we get a "symmetric" tree network, in which every aggregator is connected to roughly the same number of nodes, with high probability. Suppose now that the communication cost is proportional to the Euclidean distance between two communicating nodes. Since the number m is fixed, the communication cost in this strategy is $\Theta(n)$.

An alternative possibility (to be referred to as Design II) is to reduce the overall communication cost by using a 2-hop spanning tree. As before, we place an aggregator in each of the m sub-squares, and let the rest of the nodes in the sub-square send their messages to this aggregator. However, we allow m to be chosen optimally. The overall expected communication cost is

$$O(n/m) + O(m)$$

which we minimize by setting $m = m(n) = \Theta(n^{2/3})$, and thus reducing the $\Theta(n)$ cost of Design I to $\Theta(n^{2/3})$. On the other hand, one suspects that the detection performance of Design II

Manuscript received January 02, 2009; accepted April 07, 2009. First published May 19, 2009; current version published September 16, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Qing Zhao. This research was supported, in part, by the National Science Foundation under Grants ECCS-0701623, ANI-0335256 and ECCS-0636519, the Office of Naval Research Presidential Early Career Award for Scientists and Engineers (PECASE) N00014-09-1-0435, and the MIT Institute for Soldier Nanotechnologies. A preliminary version of this paper was presented at the Data Compression Conference, Snowbird, UT, March 2007.

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

Digital Object Identifier 10.1109/TSP.2009.2023374

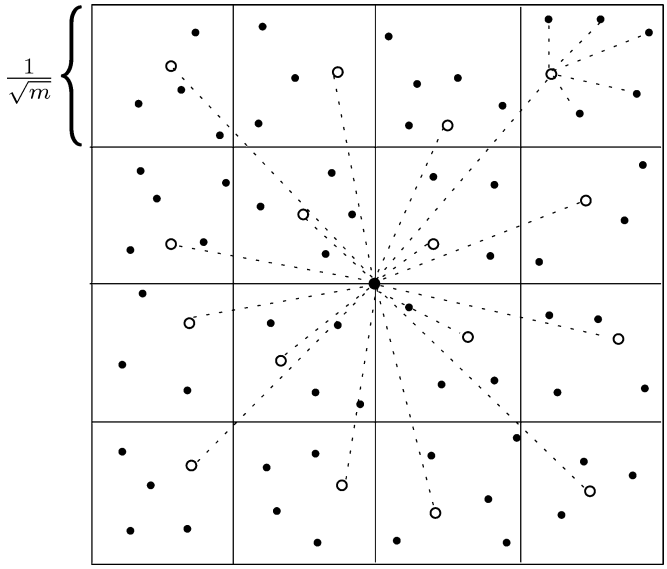

Fig. 1. Random nodes in a unit square. The hollow circles represent the local aggregators. The dotted lines represent communication links. Only one subsquare is shown with its communication links.

will be inferior to that of Design I. The results in Lemma 3 and Proposition 3 provide evidence that this is indeed the case. □

Motivated by the two designs introduced in Example 1, we will consider the detection performance of two different classes of tree networks. The first one consists of symmetric trees with a fixed number of aggregators or intermediate nodes, while the second consists of trees in which the number of intermediate nodes increases at a certain rate (we call these the rapidly branching tree sequences; cf. Section V). We characterize and compare the detection performance of these two classes, optimized over a restricted set of strategies that are easy to implement. In particular, we show in Proposition 3 that the second class performs worse than any of the tree networks in the first class.

The rest of this paper is organized as follows. In Section II, we introduce the problem formulation and some related concepts. In Section III, we show that for general tree networks, the error probability decays exponentially fast with the number of nodes in the network, and provide bounds for the rate of decay. In Sections IV and V, we consider specific classes of tree networks, characterize their performance, and provide simple (but suboptimal) strategies. Finally in Section VI, we summarize and conclude.

II. Problem Formulation

In this section, we introduce the Bayesian version of the model in [20], describe the basic assumptions and notation, and recall a useful result from [20]. We are given two hypotheses H_0 and H_1, each with prior probability $\pi_j > 0$, corresponding probability measures $\mathbb{P}_j$, and associated expectation operators $\mathbb{E}_j, j = 0, 1$. We model the sensor network as a directed rooted tree T_n, in which a node sends messages to another if there is a directed arc from the first to the second node. The root of the tree T_n is the *fusion center*, and will be denoted by f. The nodes that send messages directly to v are called its immediate *predecessors*, while v is called an immediate *successor* of each of these nodes. Let the set of immediate predecessors of a node v be $C_n(v)$.

A sequence of trees $(T_n)_{n\geq 1}$ represents the evolution of the network. We focus on tree sequences with bounded height, defined as the length of a longest directed path. For a tree with height h, a node is said to be at *level* k if it is connected to the fusion center via a path with $h-k$ hops. Hence, the fusion center f is a level h node.

We assume that under each hypothesis H_j, where $j = 0, 1$, every node v makes an i.i.d. observation X_v, with marginal distribution $\mathbb{P}_j^X$. If v is a leaf node, it sends a summary $Y_v = \gamma_v(X_v)$ of its observation to its immediate successor, where γ_v is constrained to belong to a given set Γ of allowed *quantization functions*. (For example, Γ can be the set of all binary functions of X_v.) If v is a non-leaf node, it summarizes its own observation and the messages it has received using a *transmission function* γ_v, to produce a message Y_v. This message is then sent to its immediate successor. Finally, the fusion center f uses a *fusion rule* to decide between the two hypotheses. Let Y_f be a random variable that represents the decision of the fusion center. A collection of quantization and transmission functions, one for each node, and a fusion rule will be called a *strategy*. A tree in which every non-leaf node ignores its own observation, and simply forwards a summary of its received messages, will be called a *relay tree*; in that case, non-leaf nodes will also be referred to as *relay nodes*. Let $l_n(v)$ be the number of leaves in the sub-tree rooted at node v. In particular, $l_n(f)$ is the total number of leaves of the tree T_n.

Given a tree network T_n, our objective is to minimize the probability of error $P_e(T_n) = \pi_0\mathbb{P}_0(Y_f = 1) + \pi_1\mathbb{P}_1(Y_f = 0)$, over all strategies. Let $P_e^*(T_n)$ be the minimum probability of error (over all strategies) at the fusion center. From an asymptotic perspective, we are given a sequence of trees $(T_n)_{n\geq 1}$, and seek to characterize the optimal error exponent

$$\mathcal{E}^* = \limsup_{n\to\infty} \frac{1}{n} \log P_e^*(T_n).$$

For a relay tree, we consider instead the optimal error exponent

$$\mathcal{E}_R^* = \limsup_{n\to\infty} \frac{1}{l_n(f)} \log P_e^*(T_n),$$

where we have normalized the log-error probability by $l_n(f)$, so that $\mathcal{E}_R^*$ is the error exponent per observation.

Recall that $\mathbb{P}_j^X$ is the distribution of an observation made by a node under hypothesis H_j. For any $\gamma \in \Gamma$, let $\mathbb{P}_j^\gamma$ be the distribution of $\gamma(X)$, when X has distribution $\mathbb{P}_j^X$. We make the following assumptions, which are standard in the literature (see, e.g., [5], [8], and [19]). The Kullback–Leibler (KL) divergence between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is denoted by

$$\mathrm{D}(\mathbb{P}\|\mathbb{Q}) = \mathbb{E}\left[\log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right]$$

where the expectation is taken with respect to (w.r.t.) the measure $\mathbb{P}$.

Assumption 1: The measures $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent, i.e., they are absolutely continuous w.r.t. each other. Further-

more, there exists some $\gamma \in \Gamma$ such that $-\mathrm{D}(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma) < 0 < \mathrm{D}(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma)$.

To develop insights into how the error probabilities scale with the number of nodes, we will use an upper bound for the error probabilities at each node in the network. The next proposition allows us to recursively propagate error probabilities along a tree in which *all* leaves have paths with h hops to the fusion center. Such a tree is called a *h-uniform* tree.

Let $t^{(k)} = (t_1, t_2, \ldots, t_k)$, for $k \geq 1$, and $t^{(0)} = \emptyset$. For $j = 0, 1, k \geq 1$, and $\lambda \in \mathbb{R}$, we define recursively [20]

$$\Lambda_{j,0}(\gamma;\lambda) = \Lambda_{j,0}\left(\gamma, t^{(0)};\lambda\right) = \log \mathbb{E}_j\left[\left(\frac{\mathrm{d}\mathbb{P}_1^\gamma}{\mathrm{d}\mathbb{P}_0^\gamma}\right)^\lambda\right] \quad (1)$$

$$\Lambda_{j,k}^*\left(\gamma, t^{(k)}\right) = \sup_{\lambda\in\mathbb{R}}\left\{\lambda t_k - \Lambda_{j,k-1}(\gamma, t^{(k-1)};\lambda)\right\},$$

$$\Lambda_{j,k}\left(\gamma, t^{(k)};\lambda\right) = \max\left\{-\Lambda_{1,k}^*\left(\gamma, t^{(k)}\right)(j+\lambda), \Lambda_{0,k}^*\left(\gamma, t^{(k)}\right)(j-1+\lambda)\right\}. \quad (2)$$

We make the following assumption. A prime denotes differentiation w.r.t. λ, and a double prime indicates the second derivative w.r.t. λ.

Assumption 2: Both $\mathrm{D}(\mathbb{P}_0^X||\mathbb{P}_1^X)$ and $\mathrm{D}(\mathbb{P}_1^X||\mathbb{P}_0^X)\}$ are finite, and there exists some $b \in (0,\infty)$, such that for all $\gamma \in \Gamma$, we have $\Lambda_{0,0}''(\gamma;\lambda) \leq b$ for all $\lambda \in (0,1)$, and $\Lambda_{1,0}''(\gamma;\lambda) \leq b$ for all $\lambda \in (-1,0)$.

The following result is proved as Proposition 1 in [20]. Let $S_n(v)$ be the log-likelihood ratio (or more formally, the logarithm of the associated Radon–Nikodym derivative) of the received messages at node v. A (one-bit) Log Likelihood Ratio Quantizer (LLRQ) with threshold t for a non-leaf node v is a quantizer that takes the form

$$Y_v = \begin{cases} 0, & S_n(v)/l_n(v) \leq t, \\ 1, & \text{otherwise.} \end{cases}$$

Proposition 1: Consider a sequence of h-uniform relay trees. Suppose that Assumptions 1–2 hold. Suppose that the following strategy is used: every leaf employs the same quantization function $\gamma \in \Gamma$, and every level k node $(k \geq 1)$ uses a LLRQ with threshold t_k, satisfying

$$-\mathrm{D}\left(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma\right) < 0 < \mathrm{D}\left(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma\right), \quad (3)$$

$$-\mathrm{D}\left(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma\right) < t_1 < \mathrm{D}\left(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma\right), \quad (4)$$

$$-\Lambda_{1,k-1}^*\left(\gamma, t^{(k-1)}\right) < t_k < \Lambda_{0,k-1}^*\left(\gamma, t^{(k-1)}\right), \quad \text{for } 1 < k \leq h. \quad (5)$$

Then

$$\frac{1}{l_n(f)}\log \mathbb{P}_0(Y_f = 1) \leq -\Lambda_{0,h}^*\left(\gamma, t^{(h)}\right) + \frac{n}{l_n(f)} - 1$$

$$\frac{1}{l_n(f)}\log \mathbb{P}_1(Y_f = 0) \leq -\Lambda_{1,h}^*\left(\gamma, t^{(h)}\right) + \frac{n}{l_n(f)} - 1.$$

Proposition 1 shows that the Type I and II error exponents of h-uniform trees using the strategy described in the proposition are essentially upper bounded by $-\Lambda_{0,h}^*(\gamma, t^{(h)})$ and $-\Lambda_{1,h}^*(\gamma, t^{(h)})$ respectively. In Section V, we present a class of tree networks whose error exponents are precisely $-\Lambda_{j,h}^*(\gamma, t^{(h)})$, for $j = 0, 1$.

III. Exponential Decay

In this section, we state a result that shows that the optimal error probability in a sequence of trees with bounded height h decays exponentially fast with the number of nodes n. (This is in contrast to general trees, where the decay can be sub-exponential [24].) The proof of Theorem 1 below is similar to that for the Neyman–Pearson case [20], and can be found in [25].

When $h = 1$, we have the classical parallel configuration considered in [19], and the optimal error exponent is given by

$$\mathcal{E}_P^* = -\sup_{\gamma\in\Gamma}\Lambda_{0,1}^*(\gamma, 0) = \inf_{\gamma\in\Gamma}\min_{\lambda\in[0,1]}\Lambda_{0,0}(\gamma;\lambda) < 0. \quad (6)$$

Theorem 1: Suppose that Assumptions 1 and 2 hold. Consider any sequence of trees of height h. Let $z = \liminf_{n\to\infty} l_n(f)/n$ be the asymptotic proportion of nodes that are leaves. Then

$$\mathcal{E}_P^* \leq \mathcal{E}_R^* < 0 \quad (7)$$

and

$$\min_{\lambda\in[0,1]}\log \mathbb{E}_0\left[\left(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\right)^\lambda\right] \leq \mathcal{E}^* \leq z\mathcal{E}_R^* < 0. \quad (8)$$

Furthermore, if $z = 1$, we have

$$\mathcal{E}_P^* \leq \mathcal{E}^* \leq \mathcal{E}_R^* \leq \frac{1}{2^{h-1}}\mathcal{E}_P^*. \quad (9)$$

The exact error exponent depends on several factors, such as the probability distributions and the architecture of the network. For example, in architectures that are essentially the same as the parallel configuration or can be reduced to the parallel configuration, the error exponent is $\mathcal{E}_P^*$. However, in most other cases, the error exponent is in general strictly inferior to $\mathcal{E}_P^*$ (cf. Proposition 2). To obtain some insights into the optimal error exponent, we consider specific classes of h-uniform tree networks in the next two sections. It turns out that finding optimal strategies is in general difficult, so we will instead analyze simple, but suboptimal strategies.

IV. Symmetric Tree Sequences

In this section, we consider the asymptotic performance of a special class of h-uniform tree networks, which we call r-symmetric. These are relay trees, with a bounded number of relay nodes, as in Design I in Example 1. We first characterize the optimal error exponent under a restrictive class of strategies. Then, we study the effect of the number of relay nodes on the optimal error exponent, and provide conditions under which the majority decision rule is optimal. Throughout this section, we assume that nodes can only send binary messages, taking values $\{0, 1\}$. An r-symmetric tree network is defined as follows.

Definition 1 (r-Symmetric Tree): For $h, r \geq 1$, a h-uniform tree sequence $(T_n)_{n\geq 1}$ is said to be r-symmetric if:

i) for all level k nodes v, where $k > 1$, $|C_n(v)| = r$; and
ii) for all level 1 nodes v, $l_n(v)/l_n(f) \to 1/r^{h-1}$ as $n \to \infty$.

The second condition in this definition requires that when n is large, all the r^{h-1} level 1 nodes have approximately the same number of immediate predecessors.

We define a *counting quantizer (CQ)* with threshold s for a level k node v, where $k \geq 1$, as a transmission function of the form

$$Y_v = \begin{cases} 0, & \sum_{u \in C_n(v)} Y_u \leq s \\ 1, & \text{otherwise} \end{cases}$$

where $\sum_{u \in C_n(v)} Y_u$ is the total number of 1s that v receives from its immediate predecessors. A counting quantizer has arguably the simplest possible structure. Furthermore, it is equivalent to a LLRQ with an appropriate threshold if all the messages of v's immediate predecessors are identically distributed. For tractability and to ensure that our strategies are easily implementable, we will now restrict all non-leaf nodes to using counting quantizers. We call such a strategy a *counting* strategy. Let $\mathcal{E}_S^*(r)$ denote the optimal (over all counting strategies) error exponent (in the worst-case over all r-symmetric tree sequences). We will show that with the restriction to a counting strategy, using the same transmission function at the leaves results in no loss of optimality.

For any given strategy, and for each node v, let the Type I and II error exponents be[1]

$$\psi(v) = \lim_{n\to\infty} \frac{1}{l_n(v)} \log \mathbb{P}_0(Y_v = 1)$$
$$\varphi(v) = \lim_{n\to\infty} \frac{1}{l_n(v)} \log \mathbb{P}_1(Y_v = 0).$$

Consider minimizing the following objective function:

$$\max\{\lambda_1 \psi(f),\ \lambda_2 \varphi(f)\} \tag{10}$$

where λ_1 and λ_2 are fixed positive constants. In the case of minimizing the error exponent, $\lambda_1 = \lambda_2 = 1$ [26]. We use this more general formulation because it proves to be useful later. We start with two preliminary lemmas, the first of which is proved in [19] for the case $\lambda_1 = \lambda_2$; the proof for the general case is entirely similar.

Lemma 1: Suppose that Assumptions 1–2 hold. Consider minimizing the objective function (10) at the fusion center of a parallel configuration. Then, there is no loss in optimality if we restrict all nodes to use the same transmission function, and the fusion rule to use a counting quantizer.

Consider a symmetric tree, and let the set of immediate predecessors of the fusion center f be $C_n(f) = \{v_1, \ldots, v_r\}$. From Definition 1, the subtrees rooted at the different predecessors of f are asymptotically the same. We also note that under an optimal strategy there is a tradeoff between the Type I and II error probabilities. It follows that without loss of generality, we can assume that

$$0 \geq \psi(v_1) \geq \psi(v_2) \geq \cdots \geq \psi(v_r) > -\infty \tag{11}$$
$$-\infty < \varphi(v_1) \leq \varphi(v_2) \leq \cdots \leq \varphi(v_r) \leq 0. \tag{12}$$

Furthermore, if $\psi(v_i) > \psi(v_j)$, then $\varphi(v_i) < \varphi(v_j)$, and vice versa, for all i, j.

[1]We use the notation lim here, without first showing that the limit exists. The subsequent arguments can be made completely rigorous by considering a subsequence of the tree sequence, in which limits of the Type I and II error exponents exist at each non-leaf node.

Lemma 2: To minimize the objective function (10) at the fusion center using a counting quantizer as the fusion rule, there is no loss of optimality if we restrict all immediate predecessors of f to satisfy $\psi(v_i) = \psi(v_j)$, and $\varphi(v_i) = \varphi(v_j)$ for all i, j.

Proof: Suppose the fusion center uses a counting quantizer with threshold s. Then, we have

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_f = 1) \\ &= \lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0 \left(\sum_{i=1}^{r} Y_{v_i} > s \right) \\ &= \lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_{v_i} = 1, i = 1, 2, \ldots, s+1) \\ &= \sum_{i=1}^{s+1} \lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_{v_i} = 1) \\ &= \frac{1}{r} \sum_{i=1}^{s+1} \psi(v_i) \end{aligned} \tag{13}$$

where the second equality follows because $\{Y_{v_i} = 1, i = 1, 2, \ldots, s+1\}$ is the dominating error event, and the third equality follows from independence. Similarly, we obtain

$$\lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_1(Y_f = 0) = \frac{1}{r} \sum_{i=s+1}^{r} \varphi(v_i). \tag{14}$$

Then, the objective function (10) is equal to

$$\begin{aligned} &\frac{1}{r} \max \left\{ \lambda_1 \sum_{j=1}^{s+1} \psi(v_j),\ \lambda_2 \sum_{j=s+1}^{r} \varphi(v_j) \right\} \\ &\qquad \geq \frac{1}{r} \max\{\lambda_1 (s+1) \psi(v_{s+1}),\ \lambda_2 (r-s) \varphi(v_{s+1})\} \end{aligned}$$

where equality holds if we set $\psi(v_i) = \psi(v_{s+1})$ and $\varphi(v_i) = \varphi(v_{s+1})$ for all i. Hence, it is optimal to use the same strategy for each of the sub-trees rooted at the nodes $v_1, \ldots, v_r$. ■

Theorem 2: Consider an r-symmetric tree sequence $(T_n)_{n\geq 1}$, and suppose that Assumptions 1–2 hold. Within the set of counting strategies, there is no loss in optimality if we impose the following restrictions:

i) all leaves use the same transmission function;
ii) for each $k \geq 1$, all level k nodes use counting quantizers with the same threshold.

Furthermore, the optimal error exponent at the fusion center is given by[2]

$$\begin{aligned} \mathcal{E}_S^*(r) &= \lim_{n\to\infty} \frac{1}{n} \log P_e^*(T_n) \\ &= -\sup_{\{s_k\}, t} \left\{ \left[\left(\prod_{k=2}^{h} \frac{s_k + 1}{r} \right) \Lambda_{0,1}^*(\gamma, t) \right] \right. \\ &\qquad \wedge \left[\left(\prod_{k=2}^{h} \frac{r - s_k}{r} \right) \Lambda_{1,1}^*(\gamma, t) \right] : \\ &\qquad s_k \in \{0, \ldots, r-1\}, \text{ for } k = 2, \ldots, h;\ \gamma \in \Gamma; \\ &\qquad \left. -\mathrm{D}(\mathbb{P}_0^\gamma \| \mathbb{P}_1^\gamma) < t < \mathrm{D}(\mathbb{P}_1^\gamma \| \mathbb{P}_0^\gamma) \right\}. \end{aligned} \tag{15}$$

[2]The products are taken to be 1 when $h = 1$. We also use the notation $x \wedge y = \min\{x, y\}$.

Proof (Outline[3]): From Lemma 2, we can restrict attention to counting strategies that use the same strategy at every sub-tree rooted at each $v \in C_n(f)$. Suppose that the fusion center uses, as its fusion rule, a counting quantizer with threshold s_h. Then, the objective at each $v \in C_n(f)$ is to minimize

$$\frac{1}{r}\max\left\{(s_h+1)\psi(v),\ (r-s_h)\varphi(v)\right\}.$$

We apply Lemma 2 on v, and repeat the same argument for $h-2$ steps. Therefore, we conclude that for each $k \geq 2$, there is no loss in optimality if all nodes at the same level k, use counting quantizers with the same threshold s_k. Moreover, by the same argument, there is no loss in optimality if each level 1 node has the same Type I and II error exponents. Lemma 1, applied to each level 1 node, implies that it is asymptotically optimal for all leaves to use the same transmission function γ, and all level 1 nodes to use LLRQs with the same threshold t. (Note that these LLRQs must be equivalent to counting quantizers, since the leaves use the same transmission function.) Finally, the form of the optimal error exponent is obtained by optimizing over the thresholds s_k (for $k = 2, \ldots, h$), the threshold t, and the transmission function γ. The theorem is now proved. ∎

Suppose that the transmission function γ in (15) has been fixed, and suppose that $h > 1$ and $r > 1$. Then, we have

$$\frac{1}{r^{h-1}}\prod_{i=2}^{h}(s_i+1) \leq 1$$
$$\frac{1}{r^{h-1}}\prod_{i=2}^{h}(r-s_i) \leq 1$$

and equality cannot hold simultaneously in both expressions above. Since for each $\gamma \in \Gamma$, $\Lambda^*_{0,1}(\gamma,t)$ and $\Lambda^*_{1,1}(\gamma,t)$ are continuous in t, the error exponent in (15) is achieved by setting

$$\left(\prod_{i=2}^{h}\frac{s_i+1}{r}\right)\Lambda^*_{0,1}(\gamma,t) = \left(\prod_{i=2}^{h}\frac{r-s_i}{r}\right)\Lambda^*_{1,1}(\gamma,t). \quad (16)$$

Hence, the error exponent is *strictly* smaller than that for the parallel configuration. This shows that using a r-symmetric tree results in a loss of efficiency as compared to the parallel configuration, if we restrict to counting strategies. In fact, a stronger result is possible. The detection performance of a 2-symmetric tree is strictly worse than that of a parallel configuration, even without the restriction to counting strategies.

Proposition 2: A 2-symmetric tree has strictly worse detection performance than a parallel configuration. Moreover, there is no loss in optimality restricting to counting strategies.

Proof: Consider a 2-symmetric tree with nodes v_1 and v_2 sending messages directly to the fusion center. It is not hard to see that the only choices for the fusion rule are i) declare H_0 iff both v_1 and v_2 send 0; ii) declare H_0 iff either v_1 or v_2 send a 0; iii) declare H_0 iff v_1 sends a 0; and iv) declare H_0 iff v_2 sends a 0. The latter two rules can achieve an error exponent at most half that of the parallel configuration since half of the leaves are ignored. Rules 1 and 2 are counting rules. It follows by the same argument as in the Proof of Theorem 2, that there is no loss in optimality restricting the 2-symmetric tree to counting strategies. The lemma then follows immediately from our discussion after (16). ∎

[3]For any given counting strategy, a more rigorous proof will involve taking a subsequence of $(T_n)_{n\geq 1}$ along which the vector of thresholds that defines the counting strategy converges to a limit; see the Proof of Theorem 3, for a similar argument.

A. On the Worst Case Error Exponent

When $r = 1$, the network is essentially the same, and therefore achieves the same performance, as a parallel configuration, which is the best possible. Our next result provides evidence that performance degrades as r increases. In other words, for a fixed number of nodes, it is preferable to have a high branching factor at level 1, and a low branching factor, say $r = 2$, at the other levels. Let $(T_n(r))_{n\geq 1}$ be a r-symmetric tree sequence, for $r = 1, 2, \ldots$.

Lemma 3: Suppose that Assumptions 1–2 hold, and that the network is restricted to counting strategies. Then, for any $r \geq 1, h > 1$, and any positive integer $m > 1$, $\mathcal{E}^*_S(r) < \mathcal{E}^*_S(mr)$.

Proof: Consider any sequence of integers k_i, where $i = 2, \ldots, h$, such that $0 \leq k_i < mr$ for all i. For each i, we can find an integer $s_i \in [0, r)$, such that $ms_i \leq k_i < m(s_i+1)$. Since k_i is an integer, we obtain

$$\frac{k_i+1}{mr} \leq \frac{m(s_i+1)}{mr} = \frac{s_i+1}{r} \quad (17)$$
$$1-\frac{k_i}{mr} \leq 1-\frac{ms_i}{mr} = 1-\frac{s_i}{r}. \quad (18)$$

Then, we have

$$\left[\Lambda^*_{0,1}(\gamma,t)\prod_{i=2}^{h}\frac{k_i+1}{mr}\right] \wedge \left[\Lambda^*_{1,1}(\gamma,t)\prod_{i=2}^{h}\left(1-\frac{k_i}{mr}\right)\right]$$
$$< \left[\Lambda^*_{0,1}(\gamma,t)\prod_{i=2}^{h}\frac{s_i+1}{r}\right] \wedge \left[\Lambda^*_{1,1}(\gamma,t)\prod_{i=2}^{h}\left(1-\frac{s_i}{r}\right)\right]$$
$$\leq -\mathcal{E}^*_S(r).$$

[The first strict inequality is because equality cannot hold simultaneously in both (17) and (18).] Taking the supremum over k_i, γ and t, yields $\mathcal{E}^*_S(mr) > \mathcal{E}^*_S(r)$. The proof is now complete. ∎

The above lemma shows that for any $m > 1$ and $r \geq 1$, $(\mathcal{E}^*_S(m^l r))_{l\geq 0}$ is an increasing sequence, which is bounded above by zero, hence it converges. We provide an upper bound for this limit (cf. Proposition 6) below.

Proposition 3: Suppose that Assumptions 1–2 hold. For any collection of symmetric tree sequences, $\{(T_n(r))_{n\geq 1} : r = 1, 2, \ldots\}$, where $(T_n(r))_{n\geq 1}$ is a r-symmetric tree sequence, we have

$$\limsup_{r\to\infty}\mathcal{E}^*_S(r) \leq -\sup_{\substack{\gamma\in\Gamma\\ t\in\mathbb{R}}}\left(\frac{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}}\Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}}+\Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}\right)^{h-1}.$$

Proof: Given $\gamma \in \Gamma$, and t that satisfies $-\mathrm{D}(\mathbb{P}^\gamma_0\|\mathbb{P}^\gamma_1) < t < \mathrm{D}(\mathbb{P}^\gamma_1\|\mathbb{P}^\gamma_0)$, let

$$\delta = \frac{\Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}}+\Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}, \quad (19)$$

and $s = \lfloor \delta r \rfloor$. We have

$$\mathcal{E}_S^*(r) \leq -\min\left\{\left(\frac{s+1}{r}\right)^{h-1}\Lambda_{0,1}^*(\gamma,t), \left(\frac{r-s}{r}\right)^{h-1}\Lambda_{1,1}^*(\gamma,t)\right\}.$$

Since $s/r \to \delta$ as $r \to \infty$, we obtain

$$\begin{aligned}&\limsup_{r\to\infty}\mathcal{E}_S^*(r)\\ &\quad\leq -[\delta^{h-1}\Lambda_{0,1}^*(\gamma,t)] \wedge [(1-\delta)^{h-1}\Lambda_{1,1}^*(\gamma,t)]\\ &\quad= -\left(\frac{\Lambda_{0,1}^*(\gamma,t)^{\frac{1}{h-1}}\Lambda_{1,1}^*(\gamma,t)^{\frac{1}{h-1}}}{\Lambda_{0,1}^*(\gamma,t)^{\frac{1}{h-1}}+\Lambda_{1,1}^*(\gamma,t)^{\frac{1}{h-1}}}\right)^{h-1}\end{aligned}$$

and taking the infimum over $\gamma \in \Gamma$ and $t \in \mathbb{R}$, the proposition is proved. ∎

Under some additional symmetry assumptions, the inequality in the above proposition becomes an equality. This is shown in Proposition 6 in Section V.

B. Optimality of the Majority Decision Rule

Suppose that all leaves use the transmission function $\gamma \in \Gamma$. Finding an optimal counting strategy by solving the optimization problem (15) requires us to search over a space with r^{h-1} elements, and also optimizing over t. The search can be daunting even for moderate values of r and h. For this reason, we now consider the case where r is odd, and the majority decision rule is used at every non-leaf node, i.e., a node transmits a 1 iff the majority of its immediate predecessors send a 1. For level 1 nodes, the majority decision rule corresponds to a LLRQ with threshold 0, while for nodes of level greater than 1, it corresponds to a counting quantizer with threshold $(r-1)/2$. In the proposition below, we develop a sufficient condition under which this strategy is optimal.

Proposition 4: Consider a r-symmetric tree network with $h > 1$, where r is an odd integer. Suppose that all leaves use the same transmission function γ. Let t_0 and t_1 be such that $\Lambda_{0,1}^*(\gamma,t_0) = r^{h-1}\Lambda_{1,1}^*(\gamma,t_0)$ and $\Lambda_{1,1}^*(\gamma,t_1) = r^{h-1}\Lambda_{0,1}^*(\gamma,t_1)$. Under Assumptions 1–2, and the restriction to counting strategies, if

$$\max\{\Lambda_{0,1}^*(\gamma,t_0),\Lambda_{1,1}^*(\gamma,t_1)\} \leq \frac{2r^{h-1}(r+1)\Lambda_{0,1}^*(\gamma,0)}{r^{h-1}(r-1)+r+3} \tag{20}$$

the optimal error exponent is

$$\mathcal{E}_S^*(r) = -\left(\frac{r+1}{2r}\right)^{h-1}\Lambda_{0,1}^*(\gamma,0)$$

and is achieved by using the majority decision rule at all relay nodes.

Proof: If $r = 1$, the network is equivalent to the parallel configuration, and there are no relay nodes to consider. In this case, $t_0 = t_1 = 0$ and the condition (20) holds with equality. Also, the formula for $\mathcal{E}_S^*(1)$ is the well known error exponent for the parallel configuration. Henceforth, we assume that $r > 1$.

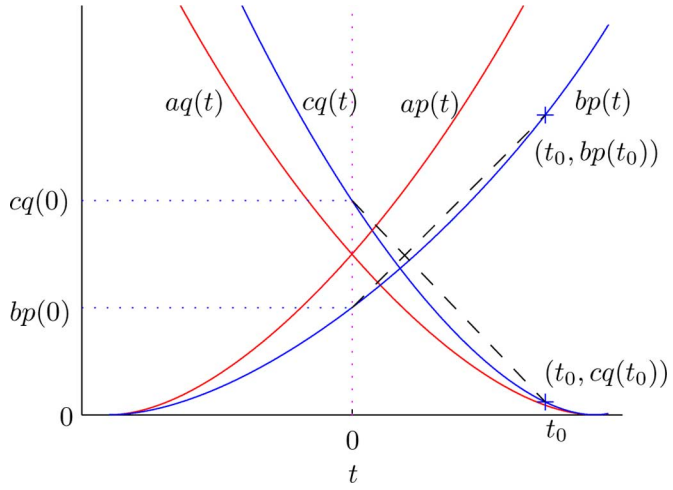

Fig. 2. Typical plot of the rate functions.

For simplicity, let $p(t) = \Lambda_{0,1}^*(\gamma,t)$ and $q(t) = \Lambda_{1,1}^*(\gamma,t)$. The sufficient condition (20) is obtained by approximating the convex functions p and q with appropriate straight line segments as shown in Fig. 2, and as we proceed to show.

Suppose that

$$b := \prod_{k=2}^{h}(s_k+1) < a := \left(\frac{r+1}{2}\right)^{h-1} < c := \prod_{k=2}^{h}(r-s_k).$$

(The argument in the case when the above inequalities hold in the reverse direction will be similar.) We consider the solution to the equations

$$\begin{aligned}y &= \frac{b(p(t_0)-p(0))}{t_0}t + bp(0)\\ y &= -\frac{c(q(0)-q(t_0))}{t_0}t + cq(0)\end{aligned}$$

which gives the intersection of the straight line approximations shown in Fig. 2. Solving the linear equations, and observing that $p(0) = q(0)$, we obtain

$$y = \frac{bc(1+d)}{c+bd}p(0)$$

where $d = (p(t_0)-p(0))/(q(0)-q(t_0))$. Since p and q are convex functions

$$\sup_t \min\{bp(t), cq(t)\} \leq y. \tag{21}$$

We first show that $y \leq ap(0)$, for all pairs (b,c) such that $b < a < c$. This is equivalent to checking that

$$d \leq \frac{c(a-b)}{b(c-a)} = \frac{a}{b}\left(1-\frac{b-a}{c-a}\right) - 1 \tag{22}$$

for all (b,c) such that $b < a < c$. Using the condition

$$\Lambda_{0,1}^*(\gamma,t_0) \leq \frac{2r^{h-1}(r+1)}{r^{h-1}(r-1)+r+3}\Lambda_{0,1}^*(\gamma,0),$$

[which is a consequence of (20)], it can be shown (after some algebra) that

$$d \leq \frac{c^*(a-b^*)}{b^*(c^*-a)}$$

where $b^* = (r+1)^{h-2}(r-1)/2^{h-1}$ and $c^* = (r+1)^{h-2}(r+3)/2^{h-1}$. The right-hand side of (22) increases when b decreases (and c increases), hence the minimum value is achieved by $b = b^*$, and $c = c^*$. This shows that (22) holds for all (b, c) such that $b < a < c$, and therefore $y \leq ap(0)$. From (21), we then have

$$\sup_{\substack{t \\ b,c:\ b\leq a\leq c}} \min\{bp(t), cq(t)\} \leq ap(0).$$

A similar argument shows that

$$\sup_{\substack{t \\ b,c:\ c\leq a\leq b}} \min\{bp(t), cq(t)\} \leq ap(0).$$

Therefore, from Theorem 2, we obtain

$$\mathcal{E}_S^*(r) \geq -\frac{a}{r^{h-1}}p(0) = -\left(\frac{r+1}{2r}\right)^{h-1}\Lambda_{0,1}^*(\gamma, 0).$$

Finally, the proposition is proved by noting that the above inequality becomes an equality when we set each of the counting quantizer thresholds to $s_k = (r-1)/2$. ■

To show that our sufficient condition in Proposition 4 is not vacuous, we provide an example in which the use of the majority decision rule does not give an optimal counting strategy.

Example 2: Consider a r-symmetric network, with $r = 45$ and $h = 3$. Suppose that each leaf sends the message 1 with probability 0.3 under hypothesis H_0, and with probability 0.9 under hypothesis H_1. If all non-leaf nodes use the majority decision rule (the counting quantizer thresholds are $s_2 = s_3 = 22$), we get an error exponent of $-129.2460/45^2$. If counting quantizers with thresholds $s_2 = s_3 = 23$ are used, our error exponent is $-129.5009/45^2$, which dominates (is more negative than) the one for the majority decision rule. In fact, it can be checked numerically that $s_2 = s_3 = 23$ is the optimal choice of counting quantizers. □

The sufficient condition in (20) can be difficult to check if one does not have access to the functions $\Lambda_{j,1}^*(\gamma, t), j = 0, 1$. A simpler but cruder sufficient condition is presented below; the proof is the same as in Proposition 4, except that we let $\mathrm{D}(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma)$ play the role of t_0, and $-\mathrm{D}(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma)$ the role of t_1.

Corollary 1: Suppose that r is an odd integer greater than 1, and that all leaves use the same transmission function γ. Under Assumptions 1–2, and the restriction to counting strategies, if

$$\max\left\{\mathrm{D}\left(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma\right), \mathrm{D}\left(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma\right)\right\} \leq -\frac{2(r+1)}{r-1}\inf_{\lambda\in[0,1]}\Lambda_{0,0}(\gamma;\lambda)$$

then using the majority decision rule at all non-leaf nodes achieves the optimal error exponent.

V. Rapidly Branching Tree Sequences

In the previous section, we considered a symmetric tree sequence in which the number of non-leaf nodes is bounded. In this section, we consider tree sequences in which the number of non-leaf nodes becomes large, in a certain sense, as n increases. We will characterize the optimal error exponent of such tree sequences under a restricted class of strategies, and show that the performance of these tree sequences is inferior to that of the r-symmetric tree sequences.

Motivated by Design II in Example 1, we define the following.

Definition 2: A rapidly branching tree sequence is a sequence of h-uniform trees $(T_n)_{n\geq 1}$, such that:

i) the number of immediate predecessors of each non-leaf node grows to infinity as n increases;

ii) there exists a sequence of positive reals $(\kappa_n)_{n\geq 1}$ such that κ_n decreases to 0 as n increases and such that for each level k node v, with $k \geq 2$, we have

$$\frac{\max_{u\in C_n(v)} l_n^2(u)}{\min_{u\in C_n(v)} l_n^2(u)} \leq \kappa_n |C_n(v)|.$$

A rapidly branching tree sequence is a sequence of trees in which the number of immediate predecessors of each node grows faster than the rate at which the tree becomes "unbalanced." The definition of a rapidly branching tree sequence implies that the number of immediate predecessors of every level 1 node grows uniformly fast, in a certain sense.

In Design II of Example 1, when n is large, with high probability, we have $l_n(u) \simeq l_n(v)$ for all level 1 nodes u and v. Therefore, this tree network fits our definition of a rapidly branching network with height $h = 2$. For a general h, a similar design can be used to approximate a h-hop MST [27]. In all of these designs, with high probability we get a rapidly branching tree network.

Since using LLRQs for every node is known to be optimal (see, e.g., [5]), we assume that every node (including leaves) is allowed to use LLRQs. The number of nodes at each level k in a rapidly branching tree network grows with n. Similar to Section IV, the problem of finding optimal LLRQs for each node in a rapidly branching tree network is, in general, intractable. Therefore, we make the following simplifying assumption.

Assumption 3: Every node is allowed to use LLRQs, and every node at the same level k uses a LLRQ with the same threshold t_k.

For notational simplicity, if each leaf uses a transmission function γ which is a LLRQ, we identify γ with the threshold of the LLRQ, i.e., $\gamma = t_0 \in \mathbb{R}$. We will first state a limit theorem for a rapidly branching tree network. This result essentially shows that the bounds in Proposition 1 are tight, and is similar in spirit to tightness results for Chernoff bounds. As the proof is rather long and tedious, we refer the reader to [25].

Proposition 5: Suppose that Assumptions 1–2 hold. Given a rapidly branching tree sequence $(T_n)_{n\geq 1}$, suppose each leaf sends its observation to its immediate successor using a transmission function $\gamma \in \Gamma$, and each level k node, where $k \geq 1$, uses a LLRQ with a common threshold t_k. Suppose that $\{\gamma, t_1, \ldots, t_h\}$ satisfy (3)–(5). Then

$$\lim_{n\to\infty}\frac{1}{l_n(f)}\log\mathbb{P}_1(Y_f = 0) = -\Lambda_{1,h}^*\left(\gamma, t^{(h)}\right)$$
$$\lim_{n\to\infty}\frac{1}{l_n(f)}\log\mathbb{P}_0(Y_f = 1) = -\Lambda_{0,h}^*\left(\gamma, t^{(h)}\right).$$

We now consider the Bayesian detection problem in a rapidly branching tree sequence, in which all nodes are constrained to sending binary messages.

Theorem 3: Consider a rapidly branching tree sequence $(T_n)_{n\geq 1}$. Suppose that Assumptions 1–3 hold. Then, the optimal error exponent is

$$\mathcal{E}^*_{RB} = -\sup_{\substack{\gamma\in\Gamma\\ t_1\in\mathbb{R}}} \left(\frac{\Lambda^*_{0,1}(\gamma,t_1)^{\frac{1}{h-1}}\Lambda^*_{1,1}(\gamma,t_1)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t_1)^{\frac{1}{h-1}}+\Lambda^*_{1,1}(\gamma,t_1)^{\frac{1}{h-1}}}\right)^{h-1}. \tag{23}$$

Furthermore, if the supremum is achieved by $\gamma\in\Gamma$, and $t_1\in(-\mathrm{D}(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma),\mathrm{D}(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma))$, then the optimal threshold for the fusion center is $t_h=0$, and the optimal threshold t_k for level k nodes, where $k=2,\ldots,h-1$, is

$$t_k = \left[\Lambda^*_{0,k-1}\left(\gamma,t^{(k-1)}\right)\Lambda^*_{1,k-1}\left(\gamma,t^{(k-1)}\right)^{\frac{1}{h-k+1}} - \Lambda^*_{0,k-1}\left(\gamma,t^{(k-1)}\right)^{\frac{1}{h-k+1}}\Lambda^*_{1,k-1}\left(\gamma,t^{(k-1)}\right)\right]\Bigg/ \left[\Lambda^*_{0,k-1}\left(\gamma,t^{(k-1)}\right)^{\frac{1}{h-k+1}}+\Lambda^*_{1,k-1}\left(\gamma,t^{(k-1)}\right)^{\frac{1}{h-k+1}}\right].$$

We first state two lemmas that we will not prove. The proof of these two lemmas are easily obtained using simple algebra.

Lemma 4: Given $k\geq 1$ and $a,b>0$, we have

$$\min_{-b<x<a}\left(\left(\frac{a+b}{a(b+x)}\right)^{\frac{1}{k}}+\left(\frac{a+b}{b(a-x)}\right)^{\frac{1}{k}}\right)^k = \left(\left(\frac{1}{a}\right)^{\frac{1}{k+1}}+\left(\frac{1}{b}\right)^{\frac{1}{k+1}}\right)^{k+1}$$

and the minimizer is given by

$$x^* = \frac{ab^{\frac{1}{k+1}}-a^{\frac{1}{k+1}}b}{a^{\frac{1}{k+1}}+b^{\frac{1}{k+1}}}.$$

Lemma 5: For $k\geq 2$, and $t^{(k)}$ satisfying (4)–(5), we have

$$\inf_{\lambda\in[0,1]}\Lambda_{0,k}\left(\gamma,t^{(k)};\lambda\right) = -\frac{\Lambda^*_{0,k}\left(\gamma,t^{(k)}\right)\Lambda^*_{1,k}\left(\gamma,t^{(k)}\right)}{\Lambda^*_{0,k}\left(\gamma,t^{(k)}\right)+\Lambda^*_{1,k}\left(\gamma,t^{(k)}\right)},$$

and

$$\Lambda^*_{1,k}\left(\gamma,t^{(k)}\right) = \frac{\Lambda^*_{1,k-1}\left(\gamma,t^{(k-1)}\right)\left(\Lambda_{0,k-1^*(\gamma,t^{(k-1)})}-t_k\right)}{\Lambda^*_{0,k-1}\left(\gamma,t^{(k-1)}\right)+\Lambda^*_{1,k-1}\left(\gamma,t^{(k-1)}\right)},$$

$$\Lambda^*_{0,k}\left(\gamma,t^{(k)}\right) = \frac{\Lambda^*_{0,k-1}\left(\gamma,t^{(k-1)}\right)\left(\Lambda^*_{1,k-1}(\gamma,t^{(k-1)}\right)+t_k)}{\Lambda^*_{0,k-1}\left(\gamma,t^{(k-1)}\right)+\Lambda^*_{1,k-1}\left(\gamma,t^{(k-1)}\right)}.$$

Proof of Theorem 3: Suppose that under Assumptions 1–3, an optimal strategy is for each leaf to use a LLRQ with threshold γ_n, and for each level k node, where $k\geq 1$, to use a LLRQ with threshold $t_{n,k}$. Let $(n_l)_{l\geq 1}$ be a subsequence such that

$$\lim_{l\to\infty}\frac{1}{l_{n_l}(f)}\log P_e(T_{n_l}) = \mathcal{E}^*_{RB}.$$

Since γ_n is bounded ($|\gamma_n|$ cannot diverge to infinity, otherwise every leaf reports either 1 or 0 with probability one asymptotically, under either hypothesis), there exists a subsequence $(u_l)_{l\geq 1}$ of $(n_l)_{l\geq 1}$ such that $\gamma_{u_l}\to\gamma\in\mathbb{R}$ as $l\to\infty$. Then, from Assumption 2, since $\mathrm{D}(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma)$ and $\mathrm{D}(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma)$ are bounded, the thresholds $t_{u_l,k}$ must satisfy $-\mathrm{D}(\mathbb{P}_0^\gamma||\mathbb{P}_1^\gamma)-1<t_{u_l,k}<\mathrm{D}(\mathbb{P}_1^\gamma||\mathbb{P}_0^\gamma)+1$, for l sufficiently large; otherwise, it can be shown that either the Type I or Type II error exponent at the fusion center is zero.

Therefore, there exists a further subsequence $(m_l)_{l\geq 1}$ of $(u_l)_{l\geq 1}$ such that for all $k, \lim_{l\to\infty}t_{m_l,k}=t_k$, for some bounded t_k. Then, for all $\epsilon>0$, from Proposition 5, we obtain

$$\mathcal{E}^*_{RB}\geq -\min\{\Lambda^*_{0,h}(\gamma+\epsilon,t_1+\epsilon,\ldots,t_k+\epsilon), \Lambda^*_{1,h}(\gamma-\epsilon,t_1-\epsilon,\ldots,t_k-\epsilon)\}.$$

Taking $\epsilon\to 0$, and noting that $\Lambda^*_{0,h}$ and $\Lambda^*_{1,h}$ are continuous in all their arguments, we get

$$\mathcal{E}^*_{RB}\geq -\min\left\{\Lambda^*_{0,h}\left(\gamma,t^{(h)}\right),\Lambda^*_{1,h}\left(\gamma,t^{(h)}\right)\right\}.$$

This shows that there is no loss in optimality if we restrict the transmission functions to be the same for all n. Therefore, it remains to optimize over $\gamma\in\Gamma$ and over $t^{(h)}$. In this case, it is well known (using the same argument as in Corollary 3.4.6 of [26]) that the optimal fusion rule at the fusion center consists of a LLRQ with threshold $t_h=0$. To simplify the notation in the following, we write $\Lambda^*_{j,k}(\gamma,t^{(k)})$ as $\Lambda^*_{j,k}$. Then, we have

$$\begin{aligned}\mathcal{E}^*_{RB} &= \inf_{\substack{\lambda\in[0,1]\\ \gamma,t^{(h-1)}}}\Lambda_{0,h-1}\left(\gamma,t^{(h-1)};\lambda\right)\\ &= -\sup_{\gamma,t^{(h-1)}}\frac{\Lambda^*_{0,h-1}\Lambda^*_{1,h-1}}{\Lambda^*_{0,h-1}+\Lambda^*_{1,h-1}} \qquad (24)\\ &= -\left[\inf_{\gamma,t^{(h-2)}}\inf_{t_{h-1}}\left\{\frac{1}{\Lambda^*_{0,h-1}}+\frac{1}{\Lambda^*_{1,h-1}}\right\}\right]^{-1}\\ &= -\left[\inf_{\gamma,t^{(h-2)}}\inf_{t_{h-1}}\left\{\frac{\Lambda^*_{0,h-2}+\Lambda^*_{1,h-2}}{\Lambda^*_{0,h-2}(\Lambda^*_{1,h-2}+t_{h-1})}+\frac{\Lambda^*_{0,h-2}+\Lambda^*_{1,h-2}}{\Lambda^*_{1,h-2}(\Lambda^*_{0,h-2}-t_{h-1})}\right\}\right]^{-1} \qquad (25)\end{aligned}$$

where (24) and (25) follow from Lemma 4. We take $a=\Lambda^*_{0,h-2}$ and $b=\Lambda^*_{1,h-2}$ in Lemma 5 to obtain

$$\mathcal{E}^*_{RB} = -\left[\inf_{\gamma,t^{(h-2)}}\left\{\left(\frac{1}{\Lambda^*_{0,h-2}}\right)^{1/2}+\left(\frac{1}{\Lambda^*_{1,h-2}}\right)^{1/2}\right\}^2\right]^{-1}.$$

The optimal error exponent and the optimal thresholds for the LLRQs then follow by repeating the above same argument for another $h-2$ steps. The proof is now complete. □

By taking $t_1 = 0$ in (23), we obtain a lower bound that matches the upper bound in (9). Hence, one does no worse than by a factor of $1/2^{h-1}$ from the optimal error exponent of a parallel configuration.

For completeness, our next result shows that the bound in Proposition 3 is an equality if leaves can use LLRQs as transmission functions. In some sense, it is also a consistency result: trees with a fixed branching factor r, in the limit of large r, perform the same as rapidly branching trees.

Proposition 6: Suppose that the set Γ of allowable transmission functions for the leaves includes LLRQs. Then, under Assumptions 1 and 2, we have

$$\lim_{r\to\infty} \mathcal{E}_S^*(r) = \mathcal{E}_{RB}^*.$$

Proof: Consider a collection of tree sequences $\{(T(n,r))_{n\geq 1} : r \geq 1\}$ such that a) each $(T(n,r))_{n\geq 1}$ is a r-symmetric tree sequence; and b) for each r and for each n, every level 1 node in $T(n,r)$ has the same number of leaves attached to it. Then, from Theorem 2, the optimal error exponent for each tree sequence $(T(n,r))_{n\geq 1}$ is $\mathcal{E}_S^*(r)$.

Suppose that there exists a subsequence $(r_m)_{m\geq 1}$ such that $g = \lim_{m\to\infty} \mathcal{E}_S^*(r_m) < \mathcal{E}_{RB}^*$. Suppose that each tree sequence $(T(n,r_m))_{n\geq 1}$ uses the asymptotically optimal counting strategy proposed in Theorem 2. Note that this strategy also satisfies Assumption 3. We shall construct a rapidly branching tree sequence from $\{(T(n,r_m))_{n\geq 1} : m \geq 1\}$. Fix a positive $\epsilon < \mathcal{E}_{RB}^* - g$, and let $(n_m)_{m\geq 1}$ be an increasing sequence of positive integers such that

$$\frac{1}{l_{n_m}(f)} \log P_e(T(n_m, r_m)) \leq \mathcal{E}_S^*(r_m) + \epsilon.$$

Let $\widetilde{T}_m = T(n_m, r_m)$. Then, it is an easy exercise to verify that $(\widetilde{T}_m)_{m\geq 1}$ satisfies Definition 2 with $\kappa_m = 1/r_m$ (which goes to 0, as $m \to \infty$). We then have

$$\begin{aligned}\frac{1}{l_{n_m}(f)} \log P_e(\widetilde{T}_m) &= \frac{1}{l_{n_m}(f)} \log P_e(T(n_m, r_m)) \\ &\leq \mathcal{E}_S^*(r_m) + \epsilon.\end{aligned}$$

Taking $m \to \infty$, we obtain

$$\limsup_{m\to\infty} \frac{1}{l_{n_m}(f)} \log P_e(\widetilde{T}_m) \leq g + \epsilon < \mathcal{E}_{RB}^*,$$

a contradiction to Theorem 3. Therefore, we must have $\liminf_{r\to\infty} \mathcal{E}_S^*(r) \geq \mathcal{E}_{RB}^*$. Finally, from Proposition 3, we obtain the desired conclusion. ■

VI. Conclusion

In this paper, we studied the detection performance of large scale tree networks with bounded height, under a Bayesian formulation. We showed that the error probability decays exponentially fast with the number of nodes in the network, and provided bounds for the rate of decay. We also considered specific classes of tree networks to quantify the detection performance. In particular, we considered simple counting strategies in symmetric tree networks, and characterized the optimal detection performance over this class of strategies. We showed that the detection performance of symmetric tree networks (with a fixed number of relay nodes) is superior to that of rapidly branching tree networks, although the latter is, in general, more energy efficient. We also showed that for these classes of tree networks and transmission strategies, the Bayesian detection performance deteriorates with the height of the tree architecture, in contrast to the results for the Neyman–Pearson formulation [20].

Throughout this paper, we have assumed that every node makes a (conditionally) i.i.d. observation. A topic for further research is the case of correlated observations, which remains a relatively unexplored area, with work mainly limited to the parallel configuration [10], [28]–[32]. ■

Acknowledgment

The authors would like to thank the anonymous reviewers for their careful reading of the manuscript, and their detailed comments that have improved the presentation.

References

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

[2] 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, 1990.

[3] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 355–361, Mar. 1992.

[4] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 550–558, 1993.

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

[6] W. W. Irving and J. N. Tsitsiklis, "Some properties of optimal thresholds in decentralized detection," *IEEE Trans. Autom. Control*, vol. 39, no. 4, pp. 835–838, 1994.

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

[8] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.

[9] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun. (Special Issue on Wireless Sensor Networks)*, vol. 22, no. 6, pp. 1007–1015, 2004.

[10] 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.

[11] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, 1988.

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

[13] 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, 1991.

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

[15] 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, 1992.

[16] 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, 1994.

[17] 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, 1995.
[18] 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.
[19] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, pp. 167–182, 1988.
[20] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155–4168, Sep. 2008.
[21] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 765–782, 1970.
[22] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, no. 3, pp. 828–835, 1969.
[23] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, 1992.
[24] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008.
[25] W. P. Tay, "Decentralized detection in resource-limited sensor network architectures," Ph.D. dissertation, Massachusetts Inst. of Technol., Cambridge, Dec. 2007.
[26] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer-Verlag, 1998.
[27] A. Clementi, M. Ianni, A. Monti, M. Lauria, G. Rossi, and R. Silvestri, "Divide and conquer is almost optimal for the bounded-hop MST problem on random Euclidean instances," in *Proc. Structural Information Communication Complexity*, Mont Saint-Michel, France, May 2005, pp. 89–98.
[28] E. Drakopoulos and C. C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 4, pp. 593–606, Jul. 1991.
[29] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 916–920, 1992.
[30] R. S. Blum and S. A. Kassam, "Optimum distributed detection of weak signals in dependent sensors," *IEEE Trans. Inf. Theory*, vol. 38, no. 3, pp. 1066–1079, May 1992.
[31] R. S. Blum, S. A. Kassam, and H. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, 1997.
[32] W. Li and H. Dai, "Distributed detection in large-scale sensor networks with correlated sensor observations," presented at the Proc. Allerton Conf. Communication, Control, Computing, Monticello, NY, Sep. 2005.

Wee Peng Tay received the B.S. degree in electrical engineering and mathematics and 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, 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, all from the Massachusetts Institute of Technology (MIT), Cambridge, in 1980, 1980, 1981, and 1984, respectively.

He is currently a Clarence J. Lebel Professor of Electrical Engineering at MIT. He also served as a Co-Director of the MIT Operations Research Center from 2002 to 2005. His research interests are in the fields of systems, optimization, communications, control, and operations research. He has coauthored four books and more than 100 journal papers in these areas.

Dr. Tsitsiklis 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 was a member of the National Council on Research and Technology in Greece from 2005 2007. His awards include an Outstanding Paper Award by the IEEE Control Systems Society in 1986, the MIT Edgerton Faculty Achievement Award in 1989, the Bodossakis Foundation Prize in 1995, and the INFORMS/CSTS prize in 1997.

Moe Z. Win (S'85–M'87–SM'97–F'04) received the B.S. degree in electrical engineering (*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 an M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering in 1998.

He is an Associate Professor at the Massachusetts Institute of Technology (MIT). Prior to joining MIT, he was at AT&T Research Laboratories for five years and at the Jet Propulsion Laboratory for seven years. His research encompasses developing fundamental theories, designing algorithms, and conducting experimentation for a broad range of real-world problems. His current research topics include location-aware networks, time-varying channels, multiple antenna systems, ultra-wide bandwidth systems, optical transmission systems, and space communications systems.

Prof. Win is an IEEE Distinguished Lecturer and IEEE Fellow, cited for "contributions to wideband wireless transmission." He was honored with the IEEE Eric E. Sumner Award (2006), an IEEE Technical Field Award for "pioneering contributions to ultra-wideband communications science and technology." Together with students and colleagues, his papers have received several awards, including the IEEE Communications Society's Guglielmo Marconi Best Paper Award (2008) and the IEEE Antennas and Propagation Society's Sergei A. Schelkunoff Transactions Prize Paper Award (2003). His other recognitions include the Laurea Honoris Causa from the University of Ferrara, Italy (2008), the Technical Recognition Award of the IEEE ComSoc Radio Communications Committee (2008), Wireless Educator of the Year Award (2007), the Fulbright Foundation Senior Scholar Lecturing and Research Fellowship (2004), the U.S. Presidential Early Career Award for Scientists and Engineers (2004), the AIAA Young Aerospace Engineer of the Year (2004), and the Office of Naval Research Young Investigator Award (2003). He has been actively involved in organizing and chairing a number of international conferences. He served as the Technical Program Chair for the IEEE Wireless Communications and Networking Conference in 2009, 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 ICC 2009 and the IEEE Semiannual International Vehicular Technology Conference in fall 2001. He was the Chair (2004–2006) and Secretary (2002–2004) for the Radio Communications Committee of the IEEE Communications Society. He is currently an Editor for 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 *Proceedings of the IEEE* (Special Issue on UWB Technology & Emerging Applications) in 2009 and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on Ultra-Wideband Radio in Multiaccess Wireless Communications) in 2002.