

On Decentralized Detection With Partial Information Sharing Among Sensors

O. Patrick Kreidl, John N. Tsitsiklis, *Fellow, IEEE*, and Spyros I. Zoumpoulis

Abstract—We study a decentralized detection architecture in which each of a set of sensors transmits a highly compressed summary of its observations (a binary message) to a fusion center, which then decides on one of two alternative hypotheses. In contrast to the star (or “parallel”) architecture considered in most of the literature, we allow a subset of the sensors to both transmit their messages to the fusion center and to also broadcast them to the remaining sensors. We focus on the following architectural question: Is there a significant performance improvement when we allow such a message broadcast? We consider the error exponent (asymptotically, in the limit of a large number of sensors) for the Neyman–Pearson formulation of the detection problem. We prove that the sharing of messages does not improve the optimal error exponent.

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

I. INTRODUCTION

WE consider a decentralized detection problem and study the value added (performance improvement) when feeding the messages (“preliminary decisions”) of some of the sensors to the remaining ones, so that the latter can take them into consideration, along with their own observations, to form their own messages. We carry out this comparison under a Neyman–Pearson formulation of the detection problem, and (for reasons of tractability) in the asymptotic regime, as the number of sensors increases. This work is part of a broader effort to understand the performance gains or losses associated with different sensor network architectures. Primarily because of analytical obstacles, this effort had been limited to the star (also called “parallel”) architecture [1]–[4] and, somewhat more generally, to tree networks [5], [6]. In contrast, the literature on architectures that allow partial information sharing among sensors is much more limited, as will be discussed shortly. The present work complements the results of [7] on a particular feedback architecture, although under somewhat restrictive assumptions.

A. Background and Related Literature

We consider a binary hypothesis testing problem, and a number of sensors each of which obtains an observation whose

distribution is determined by the true hypothesis. In a centralized system, every sensor communicates its observation to a fusion center that makes a final decision. In contrast, in a star decentralized detection architecture (introduced in the seminal work of Tenney and Sandell [1]), each sensor sends only a summary of its observation to a fusion center, in the form of a message that takes values in a finite alphabet. The fusion center then decides on one of the alternative hypotheses. The problem is to design rules through which each sensor can form its message, and through which the fusion center can interpret these messages to make a final decision, in a manner that minimizes the probability of error.

Much research has followed [1]; for a review, see [2] and [4]. For conditionally dependent observations (given the true hypothesis), the problem is NP-hard [8]. Under the assumption of conditional independence, an optimal decision rule for each sensor takes the form of a likelihood ratio test, with a suitably chosen threshold. In turn, an optimization over the set of all thresholds can yield the desired solution. Numerical algorithms for optimizing sensor thresholds in the star network have been adapted to the series, or tandem, network and later extended to any singly-rooted tree network [5]. It is now known that likelihood ratio tests remain optimal (under the conditional independence assumption) in every directed acyclic network but, with regard to tractable computation, correctness and efficiency is retained only for sparsely-connected tree-structured networks: more specifically, threshold optimization scales exponentially with the maximal degree of the nodes [9].

Similar algorithmic developments have occurred for decentralized detection architectures with feedback [10]–[12], where the fusion center may communicate a preliminary decision to the peripheral sensors for them to take into consideration when forming a next message based on a next observation. References [10] and [11] consider a Bayesian formulation at a level of generality that captures a variety of feedback architectures, allowing the sensors within each stage to exchange preliminary decisions directly (i.e., “peer communication”). Experiments show that in the presence of feedback, the fusion center’s probability of error decreases to zero as the number of feedback stages goes to infinity, and is never larger than that of the star architecture without feedback. Reference [12] considers a similarly general class of architectures but rather focuses on the use of feedback (of all sensor decisions to all sensors) in combination with successive retesting and rebroadcasting of the updated decisions to reach a consensus, an operation described as “parley.” Two modes of “parley” are examined. In the first, all sensors make a decision at each time so as to minimize the probability of error given the information available at that time; convergence to a consensus is demonstrated to occur in only a few communica-

Manuscript received July 18, 2010; accepted November 15, 2010. Date of publication December 13, 2010; date of current version March 09, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Deniz Erdogmus. This research was partially supported by the NSF under G ECCS-0701623.

The authors are with the Laboratory for Information and Decision Systems, MIT Cambridge MA 02139 USA (e-mail: opkreidl@alum.mit.edu; jnt@mit.edu; szoumpou@mit.edu).

Digital Object Identifier 10.1109/TSP.2010.2099223

tion stages at the expense of correctness (the limiting decision does not necessarily match the optimal centralized decision). In the second, the consensus decision is constrained to be correct, which requires many more stages of communication and thus comes at the expense of haste. While the sensors in these feedback architectures do remember all preceding messages from the fusion center, each sensor is forced to forget all preceding own observations; without this memory restriction, the design problem beyond the first stage of feedback essentially faces difficulties similar to those encountered in the intractable case of conditionally *dependent* observations.

The complexity of threshold optimization for large decentralized detection networks has motivated the study of a more tractable asymptotic formulation, as the number of sensors increases to infinity. Reference [13] studies the problem of optimizing the asymptotic error exponent, for the case of a star architecture with a large number of sensors that receive conditionally independent, identically distributed observations. The broader case of a network consisting of a large number of nodes arranged as a tree of bounded height is considered in [6] and [14]; the error probability is shown to decay exponentially fast with the number of nodes under both the Bayesian and the Neyman–Pearson formulations. Necessary and sufficient conditions are provided for the optimal error exponent to be the same as that corresponding to a star configuration, under the Neyman–Pearson criterion. In the same asymptotic spirit, [15] studies decentralized binary detection in a wireless sensor network where each sensor transmits its data over a multiple access channel. Under constraints on the capacity of the wireless channel, it is proven that for the problem of detecting deterministic signals in additive Gaussian noise, having a set of identical binary sensors is asymptotically optimal, as the number of observations per sensor goes to infinity.

To our knowledge, the literature on asymptotic analysis of decentralized detection networks with feedback is quite limited. Reference [16] concludes that the Neyman–Pearson performance improvement from a single stage of feedback diminishes fast with an increasing signal-to-noise ratio or an increasing number of sensors. (For the daisy-chain architecture, our results strengthen those in [16], proving no asymptotic performance gain from a single stage of feedback regardless of the signal-to-noise ratio.)

Closest to the current paper are the results of [7], which considers two architectures with feedback and studies the corresponding optimal error exponent. In one of these architectures (called FF1 in [7]), each sensor sends two bits, with the second bit formed after hearing a summary of the first message bits sent by all sensors. The authors prove that such feedback does not improve the optimal error exponent, but only under a restriction that the summary message take values in an alphabet whose size grows subexponentially with the number of sensors, thus precluding the sharing of all of the first message bits.

B. Overview

We study the Neyman–Pearson decentralized detection problem (in the asymptotic regime) for a new network architecture that we refer to as a “daisy chain.” In this architecture, the messages sent to the fusion center by the first half of the sensors

are broadcast to the second half of the sensors, before the latter get to form their own messages. This is perhaps the simplest nontrivial non-tree architecture and it features some partial feedback. While the study of non-tree or feedback architectures appears to be quite difficult in general, the daisy chain turns out to be amenable to asymptotic analysis. Indeed, we are able to prove that the additional information feedback that is present in the daisy chain does not result in a performance improvement: the optimal error exponent is the same as for the case of a star architecture with the same number of sensors.

Our study of the daisy chain parallels the work of [7] on the FF1 feedback architecture and comes to the same conclusion, namely, that the additional information sharing does not help. On the technical side, our analysis is more general than that in [7]. Besides some restrictions in [7] that can be easily removed (see Section V in [7]), that work involves a key assumption that the alphabet of the feedback messages is subexponential in the number of sensors. Furthermore, this assumption is essential to the proof in [7], which relies on the method of types. In contrast, no such assumption is present in our work. Since all of the messages of the first half of the sensors are shared, our alphabet size grows exponentially with the number of sensors. As remarked, in [7], this added generality is unnecessary if one restricts all sensors to use the same decision rule, because in that case the number of messages that equal one is a sufficient statistic that can be encoded on a much smaller alphabet. However, we do not want impose this latter restriction, because it cannot be justified *a priori*.

It is worth noting that there are certain cases where it is easily shown that feedback cannot improve asymptotic performance, for somewhat trivial reasons. One such example is binary hypothesis testing for the case of two Gaussian distributions with different means and the same variance, in the limit of a high signal-to-noise ratio. In this case, it turns out that a star architecture (with binary messages) achieves the same error exponent as a centralized architecture in which the observations are transmitted uncompressed to a fusion center [17]. By a sandwich argument, the performance of any architecture that involves binary messages plus some additional feedback falls in between, and the error exponent remains the same.

The rest of the paper is organized as follows. We formulate the decentralized detection problem for the various architectures of interest in Section II. We present the main result and its proof in Section III. Finally, we summarize and discuss possible extensions and open problems in Section IV.

II. PROBLEM FORMULATION

In this section, we introduce the classical star architecture and the daisy chain architecture. We define our notation, make the necessary probabilistic assumptions, define the performance measures of interest, and provide the necessary background.

A. Probabilistic Assumptions

We assume that the state of the environment satisfies one of two alternative hypotheses H_0 and H_1 . There is an even number, $n = 2m$, of sensors, indexed $1, \dots, n$. Each sensor i observes the realization of a random variable X_i , which takes values in an observation set \mathcal{X} , endowed with a σ -field $\mathcal{F}_{\mathcal{X}}$ of measurable

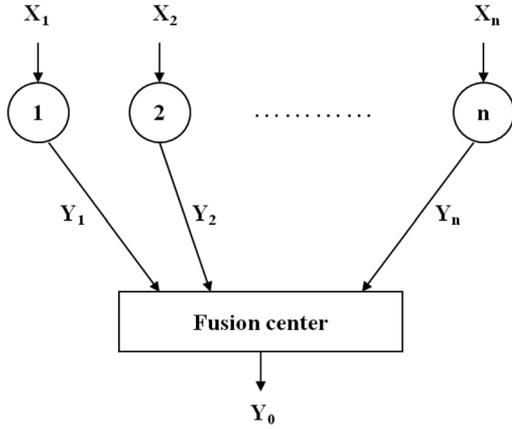


Fig. 1. The star architecture.

sets. We assume that conditioned on either hypothesis H_j , $j = 0, 1$, the random variables X_i are independent and identically distributed (i.i.d.) according to a measure \mathcal{P}_j on $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$. In the sequel we use the notation $\mathbf{E}_j[\cdot]$ to indicate an expectation taken under hypothesis H_j , and $\mathbf{P}_j(A)$ to denote the probability of an event A under H_j .

As in [13], we make the following technical assumption, which serves to facilitate the subsequent asymptotic analysis.

Assumption II.1: The measures \mathcal{P}_0 and \mathcal{P}_1 are absolutely continuous with respect to each other, but not identical. Furthermore, $\mathbf{E}_0[\log^2(d\mathcal{P}_0/d\mathcal{P}_1)] < \infty$, where $d\mathcal{P}_0/d\mathcal{P}_1$ is the Radon–Nikodym derivative of the two measures.

B. The Star Architecture

Every sensor i forms a binary message Y_i , taking values in $\{0, 1\}$, by following a rule of the form $Y_i = \gamma_i(X_i)$, for some measurable function $\gamma_i: \mathcal{X} \mapsto \{0, 1\}$. Let Γ be the set of all possible such functions. Note that a particular choice of a function $\gamma \in \Gamma$ results in particular distributions for the binary random variable $\gamma(X_1)$, under the two hypotheses. We define the Kullback–Leibler divergence $D(\gamma)$ of these two distributions by

$$D(\gamma) = \mathbf{P}_0(\gamma(X_1) = 0) \cdot \log \frac{\mathbf{P}_0(\gamma(X_1) = 0)}{\mathbf{P}_1(\gamma(X_1) = 0)} + \mathbf{P}_0(\gamma(X_1) = 1) \cdot \log \frac{\mathbf{P}_0(\gamma(X_1) = 1)}{\mathbf{P}_1(\gamma(X_1) = 1)}. \quad (1)$$

The messages Y_1, \dots, Y_n are communicated to a fusion center which uses a fusion rule of the form $\gamma_0: \{0, 1\}^n \mapsto \{0, 1\}$ and declares hypothesis H_j to be true if and only if $Y_0 = \gamma_0(Y_1, \dots, Y_n) = j$. See Fig. 1 for an illustration. Let $\Gamma_{0,n}$ be the set of all possible such functions. We use the shorthand γ^n to denote $(\gamma_1, \dots, \gamma_n)$, which is an element of Γ^n .

For a given value of n , and a particular choice of γ_0 and of $\gamma^n = (\gamma_1, \dots, \gamma_n)$, we define the probability of false alarm (Type I error) by

$$J_n^I(\gamma_0, \gamma^n) = \mathbf{P}_0(Y_0 = 1) = \mathbf{P}_0(\gamma_0(\gamma_1(X_1), \dots, \gamma_n(X_n)) = 1)$$

and the probability of missed detection (Type II error) by

$$J_n^{II}(\gamma_0, \gamma^n) = \mathbf{P}_1(Y_0 = 0) = \mathbf{P}_1(\gamma_0(\gamma_1(X_1), \dots, \gamma_n(X_n)) = 0).$$

Under a Neyman–Pearson formulation, the objective is to choose the rules γ_0 and γ^n so as to minimize the probability $\mathbf{P}_1(Y_0 = 0)$ of a missed detection subject to an upper bound α on the probability $\mathbf{P}_0(Y_0 = 1)$ of a false alarm. For our purposes, it is easier to work with the error exponent associated with the missed detection probability. We define

$$r_n(\gamma_0, \gamma^n) = \frac{1}{n} \cdot \log J_n^{II}(\gamma_0, \gamma^n) \quad (2)$$

and, for every $\alpha \in (0, 1)$,

$$R_n(\alpha) = \inf_{\gamma_0, \gamma^n} r_n(\gamma_0, \gamma^n) \quad (3)$$

where the infimum is taken over all $(\gamma_0, \gamma^n) \in \Gamma_{0,n} \times \Gamma^n$ for which $J_n^I(\gamma_0, \gamma^n) \leq \alpha$. We finally define the optimal error exponent, denoted by g_p^* , by

$$g_p^* = \lim_{n \rightarrow \infty} R_n(\alpha).$$

According to Theorem 2 in [13], the limit in the definition of g_p^* is guaranteed to exist, is the same for all values of $\alpha \in (0, 1)$, and is given by

$$g_p^* = -\sup_{\gamma \in \Gamma} D(\gamma). \quad (4)$$

Furthermore, given that the two measures \mathcal{P}_0 and \mathcal{P}_1 are not identical, it is easily seen that there exists some γ for which the distribution of $\gamma(X_1)$ is different under the two hypotheses, so that $D(\gamma) > 0$. This implies that $g_p^* < 0$.

C. The Daisy Chain Architecture

In the daisy chain architecture, the underlying probabilistic model and the sensor observations are the same as for the star architecture. What is different is that sensors $m+1, \dots, 2m = n$ get to observe the messages sent by the first m sensors before forming their own messages. We use again Y_i to denote the message sent by sensor i , and let $U = (Y_1, \dots, Y_m)$ be the additional information made available to sensors $m+1, \dots, n$. The equations that define this architecture are

$$\begin{aligned} Y_i &= \gamma_i(X_i), \quad i = 1, \dots, m \\ Y_i &= \delta_i(X_i, U), \quad i = m+1, \dots, n \\ Y_0 &= \gamma_0(Y_1, \dots, Y_n). \end{aligned}$$

For $i \leq m$, γ_i is as before the decision rule of sensor i , a measurable function from \mathcal{X} to $\{0, 1\}$. For $i > m$, the decision rule δ_i of sensor i is a measurable function from $\mathcal{X} \times \{0, 1\}^m$. Finally, the decision rule γ_0 of the fusion center is a function from $\{0, 1\}^n$ to $\{0, 1\}$. Let Γ , Δ_n , and $\Gamma_{0,n}$ be the sets of possible decision rules γ_i , δ_i , and γ_0 , respectively. We use the shorthand

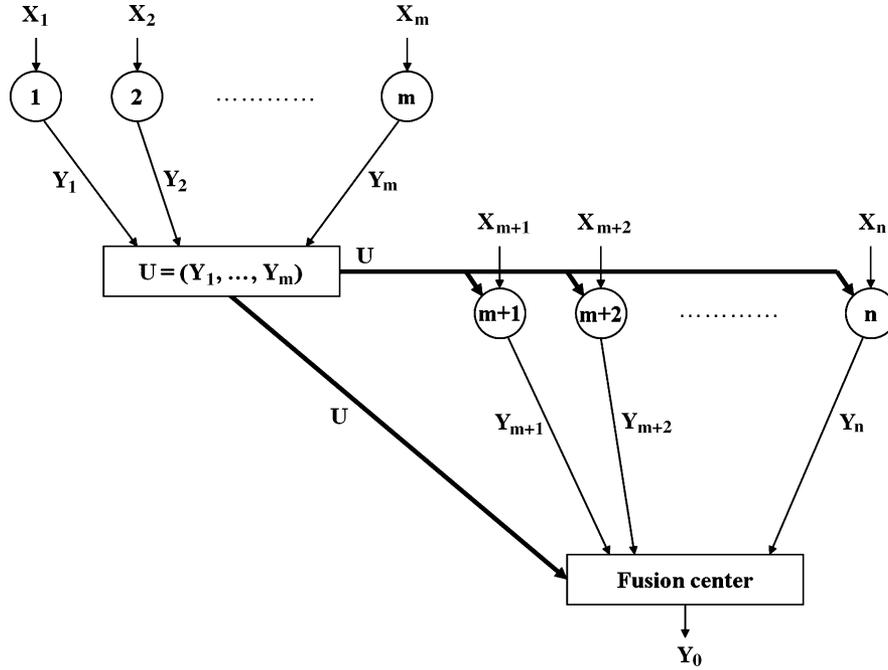


Fig. 2. The daisy chain architecture.

notation $\gamma^n = (\gamma_1, \dots, \gamma_m)$ and $\delta^n = (\delta_{m+1}, \dots, \delta_n)$, where $m = n/2$. See Fig. 2 for an illustration.

Similar to the star architecture, we define the probabilities of false alarm and missed detection associated with a collection $(\gamma_0, \gamma^n, \delta^n)$ of decision rules by

$$\begin{aligned} h_n^I(\gamma_0, \gamma^n, \delta^n) &= \mathbf{P}_0(Y_0 = 1) \\ h_n^{II}(\gamma_0, \gamma^n, \delta^n) &= \mathbf{P}_1(Y_0 = 0). \end{aligned}$$

We define

$$q_n(\gamma_0, \gamma^n, \delta^n) = \frac{1}{n} \cdot \log h_n^{II}(\gamma_0, \gamma^n, \delta^n) \quad (5)$$

and, for every $\alpha \in (0, 1)$,

$$Q_n(\alpha) = \inf_{\gamma_0, \gamma^n, \delta^n} q_n(\gamma_0, \gamma^n, \delta^n) \quad (6)$$

where the infimum is taken over all $(\gamma_0, \gamma^n, \delta^n) \in \Gamma_{0,n} \times \Gamma^n \times \Delta_n^n$ for which $h_n^I(\gamma_0, \gamma^n, \delta^n) \leq \alpha$. We finally define the optimal error exponent, denoted by $g_d^*(\alpha)$, by

$$g_d^*(\alpha) = \liminf_{n \rightarrow \infty} Q_n(\alpha).$$

It should be clear that the daisy chain architecture is “more powerful” than the star architecture: sensors $m+1, \dots, n$ are free to ignore the additional information U that they receive and emulate any possible collection of decision rules for the star architecture. For this reason, for every finite n , the optimal exponent $Q_n(\alpha)$ of the missed detection probability in the daisy chain is no larger than the optimal exponent $R_n(\alpha)$ of the missed detection probability in the star configuration. By taking the limit, it follows that

$$g_d^*(\alpha) \leq g_p^*, \quad \forall \alpha \in (0, 1).$$

For any finite n , it will generically be the case that $Q_n(\alpha) < R_n(\alpha)$, because the additional information available in the daisy chain can be exploited to some advantage. On the other hand, our main result, proved in the next section, shows that the advantage disappears in the asymptotic regime.

III. NO GAIN FROM PARTIAL INFORMATION SHARING

Our main result asserts that the optimal error exponent (as the number of sensors increases) for the daisy chain architecture is no better than that of the star configuration.

Theorem III.1: For every $\alpha \in (0, 1)$, we have $g_d^*(\alpha) = g_p^*$.

Proof: As discussed at the end of the previous section, the inequality $g_d^*(\alpha) \leq g_p^*$ is immediate. We only need to prove the reverse inequality. Toward this purpose, we consider an arbitrary choice of decision rules for the daisy chain architecture, and develop a lower bound on the probability of missed detection (for finite n), in terms of the missed detection probability for the star architecture.

Throughout we fix $\alpha \in (0, 1)$, and also an auxiliary parameter $\epsilon > 0$. Let us also fix n and decision rules $\gamma_0, \gamma^n, \delta^n$. Having done that, all of the random variables Y_i and U are well-defined. We assume that $h_n^I(\gamma_0, \gamma^n, \delta^n) = \mathbf{P}_0(Y_0 = 1) \leq \alpha$, and we will derive a lower bound on $h_n^{II}(\gamma_0, \gamma^n, \delta^n) = \mathbf{P}_1(Y_0 = 0)$.

Let¹

$$L = \log \frac{\mathbf{P}_0(U)}{\mathbf{P}_1(U)} = \sum_{i=1}^m \log \frac{\mathbf{P}_0(Y_i)}{\mathbf{P}_1(Y_i)}.$$

This is the log-likelihood ratio (a random variable) associated with the vector U of messages transmitted by the first m sensors. The first equality above is the definition of L , and the second

¹With some abuse of notation, we use $\mathbf{P}_0(U)$ to denote the random variable that takes the numerical value $\mathbf{P}_0(U = u)$ whenever $U = u$, and similarly for $\mathbf{P}_0(Y_i)$, etc.

follows from the definition $U = (Y_1, \dots, Y_m)$ and the independence of the Y_i . Let $\mu_0 = (1/m)\mathbf{E}_0[L]$. By comparing with the definition (1), and using also (4), we have that

$$\mu_0 = \frac{1}{m} \sum_{i=1}^m D(\gamma_i) \leq -g_p^*. \quad (7)$$

We say that a possible value u of the random vector U is “normal” (symbolically, $u \in N$), if

$$|L - m\mu_0| \leq \epsilon m.$$

Because of Assumption II.1, and as pointed out in [13], the (conditionally independent) random variables $\log(\mathbf{P}_0(Y_i)/\mathbf{P}_1(Y_i))$ have second moments that are bounded above (under \mathbf{P}_0) by some absolute constant c . Thus, the variance of L (under \mathbf{P}_0) is bounded above by cm . Chebyshev’s inequality then implies that

$$\mathbf{P}_0(U \notin N) \leq \frac{c}{\epsilon^2 m}.$$

We assume that m is large enough so that

$$\mathbf{P}_0(U \notin N) \leq \frac{1 - \alpha}{2(1 + \alpha)}. \quad (8)$$

Let us also say that a possible value u of the random vector U is “good” (symbolically, $u \in G$) if

$$\mathbf{P}_0(Y_0 = 1|U = u) \leq \frac{1 + \alpha}{2}.$$

We let B (for “bad”) be the complement of G .

Since $\mathbf{P}_0(Y_0 = 1) \leq \alpha$, we have

$$\begin{aligned} \alpha &\geq \mathbf{P}_0(Y_0 = 1) \\ &= \sum_u \mathbf{P}_0(U = u) \mathbf{P}_0(Y_0 = 1|U = u) \\ &\geq \sum_{u \in B} \mathbf{P}_0(U = u) \mathbf{P}_0(Y_0 = 1|U = u) \\ &\geq \frac{1 + \alpha}{2} \cdot \sum_{u \in B} \mathbf{P}_0(U = u) \\ &= \frac{1 + \alpha}{2} \cdot \mathbf{P}_0(U \in B). \end{aligned}$$

Thus,

$$\mathbf{P}_0(U \notin G) = \mathbf{P}_0(U \in B) \leq \frac{2\alpha}{1 + \alpha}. \quad (9)$$

Using (8) and (9), we obtain

$$\begin{aligned} \mathbf{P}_0(U \in N \cap G) &\geq 1 - \mathbf{P}_0(U \notin N) - \mathbf{P}_0(U \notin G) \\ &\geq 1 - \frac{1 - \alpha}{2(1 + \alpha)} - \frac{2\alpha}{1 + \alpha} = \frac{1 - \alpha}{2(1 + \alpha)} \\ &> 0. \end{aligned} \quad (10)$$

We will now argue that conditioned on $U = u$, and for $u \in G$, the missed detection probability admits an $\Omega(e^{mg_p^*})$ lower bound. Suppose that a certain value u of the random vector U has been realized. Conditioned on this event, and for this given value of u , the final decision is determined by a rule of the form

$$Y_0 = \gamma_0(u, \delta_{m+1}(X_{m+1}, u), \dots, \delta_n(X_n, u)).$$

Since u has been fixed to a constant, this is of the form

$$Y_0 = \bar{\gamma}_0(\bar{\delta}_{m+1}(X_{m+1}), \dots, \bar{\delta}_n(X_n)),$$

for suitable functions $\bar{\gamma}_0$ and $\bar{\delta}_i$. (Of course these functions depend on the specific choice of u .) We recognize this as the expression for Y_0 in a decentralized detection problem with a star architecture and m sensors.

For the constant value of u under consideration, and since $u \in G$, the false alarm probability is bounded above by $\alpha' = (1 + \alpha)/2 < 1$. We now invoke the definition of g_p^* , suitably translated to the case of a finite number of sensors. It implies that for the given $\alpha' \in (0, 1)$, there exists some n_0 (depending on ϵ and α') such that if $J_m^I(\gamma_0, \gamma^m) \leq \alpha'$, then

$$J_m^{II}(\gamma_0, \gamma^m) \geq e^{mg_p^* - m\epsilon}, \quad \forall m \geq n_0$$

where m is the number of sensors in the star architecture. By applying this observation to the last m sensors of the daisy chain architecture, and conditioned on $U = u$, we obtain, for $m \geq n_0$,

$$\mathbf{P}_1(Y_0 = 0|U = u) \geq e^{mg_p^* - m\epsilon}, \quad \forall u \in G. \quad (11)$$

Let us now suppose that $u \in N$. From the definition of the log-likelihood ratio L , we have

$$\mathbf{P}_1(U) = e^{-L} \mathbf{P}_0(U).$$

When $U \in N$, we also have $L \leq m\mu_0 + m\epsilon$, and using also (7), we obtain

$$\begin{aligned} \mathbf{P}_1(U = u) &\geq e^{-m\mu_0 - m\epsilon} \mathbf{P}_0(U = u) \\ &\geq e^{mg_p^* - m\epsilon} \mathbf{P}_0(U = u), \quad \text{if } u \in N. \end{aligned} \quad (12)$$

We now apply the usual change of measure argument. We have, for $m \geq n_0$,

$$\begin{aligned} \mathbf{P}_1(Y_0 = 0) &= \sum_u \mathbf{P}_1(U = u) \mathbf{P}_1(Y_0 = 0|U = u) \\ &\geq \sum_{u \in N \cap G} \mathbf{P}_1(U = u) \mathbf{P}_1(Y_0 = 0|U = u) \\ &\geq e^{mg_p^* - m\epsilon} \sum_{u \in N \cap G} \mathbf{P}_0(U = u) \mathbf{P}_1(Y_0 = 0|U = u) \\ &\geq e^{mg_p^* - m\epsilon} \sum_{u \in N \cap G} \mathbf{P}_0(U = u) e^{mg_p^* - m\epsilon} \\ &= e^{ng_p^* - n\epsilon} \mathbf{P}_0(U \in N \cap G) \\ &\geq e^{ng_p^* - n\epsilon} \frac{1 - \alpha}{2(1 + \alpha)}. \end{aligned}$$

Here, the second inequality follows from (12), and the third inequality from (11). The next equality is because $n = 2m$. The last inequality follows from (10).

Taking logarithms, dividing by n , and taking the limit as $n \rightarrow \infty$, we obtain that $g_d^*(\alpha) \geq g_p^* - \epsilon$. Since ϵ was arbitrary, the desired result follows. ■

IV. DISCUSSION, EXTENSIONS, AND OPEN PROBLEMS

We have proved that the daisy chain architecture introduced and analyzed in this paper performs no better (in the sense of the

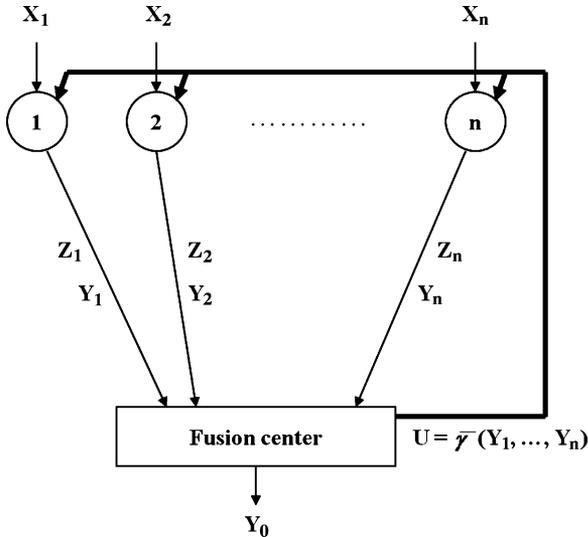


Fig. 3. An architecture with more feedback.

asymptotically optimal Neyman–Pearson error exponent) than a star architecture with the same number of sensors and observations. This is despite the fact that the daisy chain architecture provides substantially richer information to the last half of the sensors.

Our main result was given for the case where the number of sensors in the first and the second stage of the daisy chain architecture is the same (and grows with the total number of sensors in the architecture). We note that the proof can easily be modified to show the same result for the case where the numbers of sensors in the first and second stages are different, as long as they both grow to infinity.

This work opens up a number of research questions. One involves the case of a Bayesian performance criterion: we are given prior probabilities for the two hypotheses (which do not change with the number of sensors) and wish to minimize the overall probability of error. From earlier works [6], [14], we know that Bayesian formulations can be qualitatively different. In particular, while for certain classes of trees the Neyman–Pearson error exponent is the same for tree and star architectures, this is not necessarily the case for the Bayesian error exponent. Determining whether feedback provides some added value (in terms of the optimal Bayesian error exponent) is an open problem for the daisy chain as well as for the other architectures to be discussed below.

Let us now consider a few related architectures. The daisy chain architecture that we studied was “information-rich,” in the sense that *all* messages from the first m sensors were transmitted to the remaining sensors. We can make this architecture less rich, by reducing the amount of shared information. The most information-deprived architecture with a daisy chain structure that we can think of is described next; we refer to it as the “minimal daisy chain” architecture. Here, the messages $Y_i = \gamma_i(X_i)$ of the first m sensors are sent to a fusion center, as before. This fusion center computes a binary preliminary decision $U = \bar{\gamma}(Y_1, \dots, Y_m)$, which it then transmits to the sensors $m + 1, \dots, n$. The latter sensors send messages of

the form $Y_i = \delta_i(X_i, U)$ to a fusion center. We further restrict the fusion center to “forget” the messages it received from the first m sensors, and require it to make a final decision of the form $Y_0 = \gamma_0(Y_{m+1}, \dots, Y_n)$. Because the sensors and the fusion center have access to less information, the optimal (Neyman–Pearson) error exponent of this new architecture (denoted $g_{md}^*(\alpha)$) cannot be better, so that $g_p^* = g_d^* \leq g_{md}^*(\alpha)$. On the other hand, this architecture is at least as powerful as the tree architecture considered in Example 1 of [6], whose error exponent is known to be equal to g_p^* . (This is the tree architecture in which U is only sent to sensor $m + 1$, who then forwards it to the final fusion center.) This establishes that $g_{md}^*(\alpha) = g_d^*$, for all $\alpha \in (0, 1)$. For some additional results, [17] shows that for the minimal daisy chain architecture, there is no loss of optimality if sensors $1, \dots, m$ all use the same decision rule, and all but one of the sensors $m + 1, \dots, n$ use the same decision rule. (The analogous result, that there is no loss of optimality in the “information-rich” daisy chain architecture if sensors $1, \dots, m$ all use the same decision rule, and if sensors $m + 1, \dots, n$ all use the same decision rule, follows trivially from the fact that sensors $m + 1, \dots, n$ are free to ignore the additional information U that they receive and emulate any possible collection of decision rules for the star architecture.)

An alternative architecture that incorporates more feedback and which raises interesting research questions is the following. Each of n sensors sends a binary message $Y_i = \gamma_i(X_i)$ to a fusion center. A summary $U = \bar{\gamma}(Y_1, \dots, Y_n)$ of these messages is formed. One can of course discuss many variants. In one of them, U is restricted to be binary; in another, we have $U = (Y_1, \dots, Y_n)$. Then, each of the n sensors gets to transmit a second binary message Z_i , according to a rule $Z_i = \delta_i(X_i, U)$. Finally, a fusion center makes a decision Y_0 , according to a rule of the form $Y_0 = \gamma_0(Y_1, \dots, Y_n, Z_1, \dots, Z_n)$. This architecture is shown in Fig. 3. The optimal (Neyman–Pearson) error exponent of any variant is to be compared with the error exponent for a star architecture in which each sensor gets to transmit two binary messages, based only on its own information. (This star architecture is essentially the same model as above, except that now Z_i is given by a rule of the form $Z_i = \delta_i(X_i)$.) As discussed earlier, [7] proves that there is no gain if U takes values in an alphabet whose size grows at most subexponentially with n (as in the first variant proposed above). However, for the case of exponential alphabets (as in the second variant discussed above), it is unknown at present whether the error exponent matches the error exponent for a star architecture in which each sensor gets to transmit two binary messages, or whether such feedback can improve the error exponent.

More generally, the range of possible non-tree or feedback architectures is vast. While such architectures tend to lead into intractable problems (as far as the optimal exponent is concerned), it may be possible to identify some that are tractable or to carry out some qualitative comparisons.

ACKNOWLEDGMENT

The authors would like to thank Prof. W. Peng Tay and the anonymous reviewers for their feedback on the original manuscript and for bringing reference [7] to our attention.

REFERENCES

- [1] R. R. Tenney and J. N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, 1981.
- [2] J. N. Tsitsiklis, "Decentralized detection," *Adv. Stat. Signal Process.*, vol. 2, pp. 297–344, 1993.
- [3] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.
- [4] J.-F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, May 2007.
- [5] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Tree structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 1, pp. 211–231, 1993.
- [6] 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.
- [7] H. M. H. Shalaby and A. Papamarcou, "A note on the asymptotics of distributed detection with feedback," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 633–640, Mar. 1992.
- [8] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, no. 3, pp. 440–446, 1985.
- [9] O. P. Kreidl and A. S. Willisky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Trans. Autom. Control*, vol. 55, pp. 563–578, 2010.
- [10] 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.
- [11] S. Alhakeem and P. K. Varshney, "Decentralized Bayesian detection with feedback," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 4, pp. 503–513, Jul. 1996.
- [12] P. F. Swaszek and P. Willett, "Parley as an approach to distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 447–457, Jan. 1995.
- [13] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signal, Syst.*, vol. 1, pp. 167–182, 1988.
- [14] W. P. Tay, "Decentralized detection in resource-limited sensor network architectures," Ph.D. dissertation, Massachusetts Inst. of Technology, Cambridge, MA, Feb. 2008.
- [15] 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.
- [16] D. A. Pantos, K. W. Halford, D. Kazakos, and P. Papantoni-Kazakos, "Distributed binary hypothesis testing with feedback," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 1, pp. 21–42, Jan. 1995.
- [17] S. I. Zoumpoulis, "Decentralized detection in sensor network architectures with feedback," M.Eng. dissertation, Massachusetts Inst. of Technology, Cambridge, MA, May 2009.



O. Patrick Kreidl received the B.S. degree from George Mason University (GMU), Fairfax, VA, in 1994 and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2008, respectively, all in electrical engineering.

He has been a Research Affiliate at MIT's Laboratory for Information and Decision Systems since 2008, and a Research Engineer at BAE Systems Advanced Information Technologies (formerly Alphatech, Inc.) since 1996. He also held an adjunct faculty appointment within GMU's Electrical and Computer Engineering Department from 1998 to 2001. His research interests are in statistical signal processing, stochastic control, and numerical optimization.



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 served as a Co-Director of the MIT Operations Research Center from 2002 to 2005, and in the National Council on Research and Technology in Greece (2005 to 2007). His research interests are in systems, optimization, communications, control, and operations research. He has coauthored four books and more than 100 journal papers in these areas.

Dr. Tsitsiklis is a member of the National Academy of Engineering and holds an honorary doctorate from the Université Catholique de Louvain.



Spyros I. Zoumpoulis received the B.S. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2008 and 2009, respectively. He is currently working towards the Ph.D. degree at the Department of Electrical Engineering and Computer Science, MIT.

His research interests are in signal processing, economics, and operations research.