Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Distributed Hypothesis Testing with a Fusion Center: The Conditionally Dependent Case

Kien C. Nguyen, Tansu Alpcan, and Tamer Başar

***Abstract*— The paper deals with decentralized Bayesian detection with M hypotheses, and N sensors making conditionally correlated measurements regarding these hypotheses. Each sensor sends to a fusion center an integer from $\{0, 1, .., D-1\}$, and the fusion center makes a decision on the actual hypothesis based on the messages it receives from the sensors so as to minimize the average probability of error. Such conditionally dependent scenarios arise in several applications of decentralized detection such as sensor networks and network security. Conditional dependence leads to a non-standard distributed decision problem where threshold based policies (on likelihood ratios) are no longer optimal, which results in a challenging distributed optimization/decision making problem. We show that, in this case, the minimum average probability of error cannot be expressed as a function of the marginal distributions of the sensor messages. Instead, we characterize this probability based on the joint distributions of these messages. We also provide some numerical results for the case where the sensors' measurements follow bivariate normal distributions.**

I. INTRODUCTION

Centralized hypothesis testing has been examined in many papers and texts (see, for example, [1]). Tenney and Sandell [2] were the first to study hypothesis testing within a decentralized setting, where each of two sensors locally selected its threshold for the likelihood ratio test to minimize a common cost function. Sadjadi [3] later extended this work to accommodate arbitrary numbers of sensors and hypotheses. The paper did not consider a fusion center: the cost was a function of the sensor decisions and the true hypothesis. A comprehensive survey of decentralized detection can be found in [4], which examined different decentralized detection structures with both conditionally independent and correlated sensor observations. The complexity of decentralized detection problems was also studied in [5]. In [6], Hoballah and Varshney proposed a Person-By-Person Optimization (PBPO) scheme to optimize a distributed detection system using the Bayesian criterion. The decentralized detection problem with quantized observations was addressed in [7], where the authors also introduced a joint power constraint on the sensors. An extension to [7] was given in [8], where the constraint was placed on the average cost of the system.

For a single sensor, it has been proved in [9] that the set of conditional distributions, Q, is a compact set, and thus any cost function that is a continuous function on Q will attain a minimum, which corresponds to an optimal quantizer. In a parallel configuration with multiple sensors and a fusion center, if the sensor observations are independent given each hypothesis, it has also been shown in [9] that there exists an optimal solution over the Cartesian product of the sets of conditional marginal probabilities $\{P_i(d)\}, i \in \{0, 1, \dots, M-1\}, d \in \{0, 1, \dots, D-1\}$.

However, in several applications of hypothesis testing such as sensor networks and attack/anomaly detection, it is generally seen that the observations from different sensors may be correlated (see, for example, [10], [11], [12], [13]). It is this scenario we address in this paper. We show that when the observations are conditionally dependent, minimum average probability of error, P_e, can no longer be expressed as a function of the marginal probabilities. We then proceed to characterize P_e based on the set of joint probabilities of the sensor messages. We show that there exist optimal solutions for both the general case and the special case where the sensors are restricted to threshold rules based on likelihood ratios.

The paper is organized as follows. In Section II, we formulate the problem and specify the decision rules of the sensors and the fusion rule of the fusion center. Next, in Section III, we derive the relationships among the minimum probability of error, the marginal distributions, and the joint distributions of sensor messages, given that the sensor observations are conditionally correlated. We provide an example where the joint distributions of the sensor observations are bivariate normal in Section IV. Finally, some concluding remarks end the paper.

II. PROBLEM FORMULATION

A. Background

We consider the decentralized Bayesian detection problem with a parallel configuration, where N sensors are directly connected to a fusion center. The sensors observe M hypotheses ($M \geq 2$), $H_0, H_1 \dots, H_{M-1}$, whose prior probabilities $\pi_0, \pi_1 \dots, \pi_{M-1}$ are known. The observations of the sensors are $Y_1, Y_2, \dots, Y_N$, where Y_j is a random variable that takes values in an appropriately defined finite or infinite set $\mathcal{Y}_j, j = 1, \dots, N$. Given hypothesis H_i, the joint distribution of the observations is $P_i(y_1, \dots, y_N)$, where $i = 0, 1, \dots, M-1$. Sensor observations are not assumed to be conditionally independent nor identically distributed. Each sensor uses a decision rule, which is a map $\gamma_j : \mathcal{Y}_j \mapsto \{0, 1, \dots, D-1\}$, and then sends the resulting message,

This work was supported by Deutsche Telekom Laboratories and in part by the Vietnam Education Foundation.

Tamer Başar and Kien C. Nguyen are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, 1308 W Main St., Urbana, IL 61801, USA basar1@illinois.edu, knguyen4@illinois.edu

Tansu Alpcan is with Deutsche Telekom Laboratories, Ernst-Reuter-Platz 7, D-10587 Berlin, Germany tansu.alpcan@telekom.de

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

which is an integer $d_j \in \{0, 1, \ldots, D-1\}$, to the fusion center. We take the communication channels between the sensors and the fusion center to be perfect. At the fusion center, a fusion rule $\gamma_0 : \{0, 1, \ldots, D-1\}^N \mapsto \{0, 1 \ldots, M-1\}$ is employed to finally decide which hypothesis is true. Using the Bayesian approach, we seek a joint optimization of the decision rules at the sensors and the fusion rule to minimize the probability of error P_e at the fusion center. The configuration of the N sensors and the fusion center are shown in Figure 1.

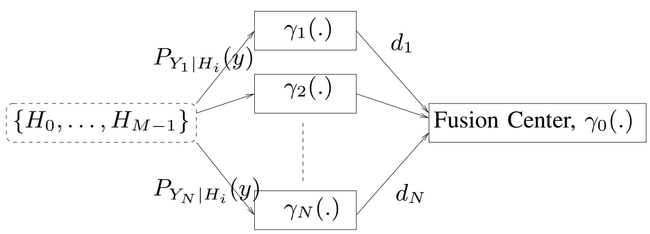

Fig. 1. Decentralized hypothesis testing with N sensors and a fusion center.

Taking a realization of the random variable Y_j and sending out a message in $\{0, 1, \ldots, D-1\}$, each sensor can be considered as a quantizer. As mentioned in the Introduction, [9] characterizes these quantizers based on the set of marginal distributions of the messages given each hypothesis. Following [9], let

$$\begin{aligned} q_d(\gamma_j|H_i) &= Pr(\gamma_j(Y_j) = d|H_i),\ i = 0, \ldots, M-1, \\ j &= 1, \ldots, N,\ d = 0, \ldots, D-1. \end{aligned} \tag{1}$$

For any $\gamma_j \in \Gamma_j$, where Γ_j is the set of all deterministic quantizers for sensor j, let

$$q(\gamma_j|H_i) = (q_0(\gamma_j|H_i), \ldots, q_{D-1}(\gamma_j|H_i)). \tag{2}$$

Define the vector $q(\gamma_j) \in R^{MD}$, for any $\gamma_j \in \Gamma_j$, as

$$q(\gamma_j) = (q(\gamma_j|H_0), \ldots, q(\gamma_j|H_{M-1})). \tag{3}$$

Now a quantizer can be represented by its vector $q(\gamma)$ for the purpose of detecting the hypotheses. Let

$$Q_j = \{q(\gamma_j) : \gamma_j \in \Gamma_j\}. \tag{4}$$

For a parallel configuration with N sensors, we define

$$q(\gamma_1, \gamma_2, \ldots, \gamma_N) = (q(\gamma_1), q(\gamma_2), \ldots, q(\gamma_N)). \tag{5}$$

Then we have $q(\gamma_1, \gamma_2, \ldots, \gamma_N) \in Q_a$, where Q_a is the Cartesian product of all $Q_j, j = 1, \ldots, N$: $Q_a = \times_{j=1}^N Q_j$.

As previously mentioned, it has been proved in [9] that Q_j is a compact set, and thus any cost function that is a continuous function on Q_j will attain a minimum. In a parallel configuration with multiple sensors and a fusion center, if the sensor observations are independent given each hypothesis, it has also been shown that there exists an optimal solution over the set Q_a [9].

B. Decision Rules at the Sensors and the Fusion Center

First we define two classes of decision rules at each sensor and the fusion center. (A fusion center can also be viewed as a sensor; thus we use the term "sensor" to refer to both in this subsection.) A *general rule* is one in which the observation space is partitioned into M regions, $R_i, i = 0, 1, \ldots, M-1$, and the sensor will pick H_i if $Y \in R_i$. In the scope of this paper, we define the *threshold rule* for the case of binary hypotheses ($M = 2$) as follows. A threshold rule is a general rule where

$$R_1 = \left\{ y \in \mathcal{Y} : \frac{P_1(y)}{P_0(y)} \geq \tau \right\} \tag{6}$$

$$R_0 = \left\{ y \in \mathcal{Y} : \frac{P_1(y)}{P_0(y)} < \tau \right\} \tag{7}$$

where $\mathcal{Y}$ is the observation space of the sensor, and $P_0(y)$ and $P_1(y)$ are the conditional distributions of the observation given H_0 and H_1, respectively.

Assuming uniform costs, the Bayes risk will become the average probability of error [1]. As mentioned above, the fusion center can be considered as a sensor with the observation being $(d_1, d_2, \ldots, d_N)$. Note that we seek a joint optimization of the decision rules at the (local) sensors and the fusion rules at the fusion center to minimize the Bayes risk. However, if the decision rules at the (local) sensors have already been optimized, the fusion rule at the fusion center must be the solution to the centralized detection problem to minimize the Bayes risk. From [1], the fusion rule for binary hypotheses can be written as a likelihood ratio test:

$$\gamma_0(d_1, d_2, \ldots, d_N) = \begin{cases} 1 & \text{if } \frac{P_1(d_1, d_2, \ldots, d_N)}{P_0(d_1, d_2, \ldots, d_N)} \geq \frac{\pi_0}{\pi_1} \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

and the corresponding average probability of error at the fusion center is given as

$$\begin{aligned} P_e &= \pi_0 P_0 \left(L_a \geq \frac{\pi_0}{\pi_1} \right) + \pi_1 P_1 \left(L_a < \frac{\pi_0}{\pi_1} \right) \\ &= \pi_0 \sum_{(d_1, d_2, \ldots, d_N): L_a \geq \frac{\pi_0}{\pi_1}} P_0(d_1, d_2, \ldots, d_N) \\ &\quad + \pi_1 \sum_{(d_1, d_2, \ldots, d_N): L_a < \frac{\pi_0}{\pi_1}} P_1(d_1, d_2, \ldots, d_N) \\ &\quad \text{where } L_a = \frac{P_1(d_1, d_2, \ldots, d_N)}{P_0(d_1, d_2, \ldots, d_N)}. \end{aligned} \tag{9}$$

Here $P_i(d_1, d_2, \ldots, d_N)$, $i = 0, 1$, are the conditional joint probability density functions (given H_i) of the sensor messages, which can be computed as follows

$$P_i(d_1, d_2, \ldots, d_N) = \int_{R_{d_N}^{(N)}} \cdots \int_{R_{d_1}^{(1)}} P_i(y_1, \ldots, y_N) \, dy_1 \ldots dy_N \tag{10}$$

where $d_j = 0, 1, \ldots, D-1$ and $R_{d_j}^{(j)}$ is the region where sensor j decides to send message d_j, $j = 1, \ldots, N$. Thus, it can be seen that in the optimal solution (which achieves the minimum P_e) the fusion rule is always a likelihood ratio test (8), but the decision rules at the local sensors can be general rules. It has been shown in [4] that when the sensor observations are independent given each hypothesis, the optimal solution can be achieved with the decision rule at each sensor being also a threshold rule. However, when the sensor observations are conditionally dependent, the threshold rules at the local sensors can no longer achieve the minimum P_e in general [4]. It is also worth noting that, in general, the minimum P_e at the fusion center only depends on the decision rules at the sensors. If we restrict the sensors to threshold rules, the minimum P_e will only depend on the thresholds at the sensors, $\{\tau_1, \tau_2, \ldots, \tau_N\}$.

III. THE EXISTENCE OF OPTIMAL SOLUTIONS

In this section, we first prove that when the observations are conditionally dependent, P_e can no longer be expressed as a function of the marginal distributions of the messages from the sensors. We then characterize P_e based on the set of joint distributions of the sensor messages. We show that this set is compact and there exists an optimal solution (that minimizes P_e) when general rules are used at the sensors, and there also exists an optimal solution when the sensors are restricted to threshold rules. Propositions 1 and 2 are stated for $D = 2$ and $M = 2$ but their results can be extended to $M > 2$.

Proposition 1: Let $f_0(y_1, y_2)$ and $f_1(y_1, y_2)$ be two non-identical joint probability density functions, where $f_i(y_1, y_2)$, $i = 0, 1$, is continuous on $\mathcal{R}^2$ and nonzero for $-\infty < y_1, y_2 < \infty$. Let $\Phi_i(y_1, y_2)$, $i = 0, 1$, denote the corresponding cumulative distribution functions. Let

$$\alpha_0 = \Phi_0(y_1^*, y_2^*) = \int_{-\infty}^{y_1^*}\int_{-\infty}^{y_2^*} f_0(y_1, y_2)dy_2dy_1, \quad (11)$$

$$\alpha_1 = \Phi_1(y_1^*, y_2^*) = \int_{-\infty}^{y_1^*}\int_{-\infty}^{y_2^*} f_1(y_1, y_2)dy_2dy_1. \quad (12)$$

where (y_1^*, y_2^*) is an arbitrary point in $\mathcal{R}^2$. Then, specifying a value for $\alpha_0 \in (0, 1)$ does not uniquely determine the value of α_1, and vice versa.

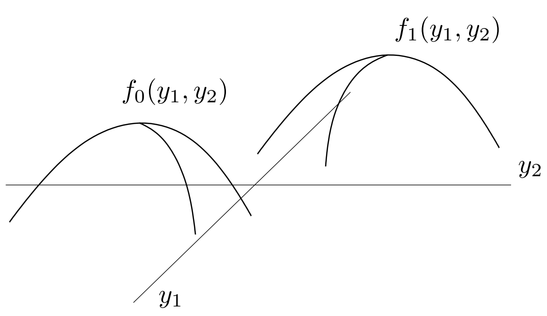

Fig. 2. α_0 and α_1 are integrations of $f_0(y_1, y_2)$ and $f_1(y_1, y_2)$ over the same region.

Proof: Let $g_i(y_1)$ and $h_i(y_2)$ be the marginal densities of y_1 and y_2 given H_i, where $i = 0, 1$. For each $0 < \alpha_0 < 1$, we can pick $\gamma_0 > 0$ such that $\alpha_0 + \gamma_0 < 1$. As the conditional marginal density $g_0(y_1)$ is continuous, we can always uniquely pick y_1^* such that $\int_{-\infty}^{y_1^*} g_0(y_1)dy_1 = \alpha_0 + \gamma_0$. Once y_1^* is specified, we can also choose y_2^* such that $\int_{-\infty}^{y_1^*}\int_{-\infty}^{y_2^*} f_0(y_1, y_2)dy_2dy_1 = \alpha_0$. Thus, for each fixed value of γ_0, we have a unique pair (y_1^*, y_2^*). It can be seen that there are infinitely many values of γ_0 satisfying $\alpha_0 + \gamma_0 < 1$, each of which yields a different pair (y_1^*, y_2^*). Therefore, specifying a value for $\alpha_0 \in (0, 1)$ does not uniquely determine the value of α_1, and vice versa, unless $f_0(y_1, y_2)$ and $f_1(y_1, y_2)$ are identically equal. ■

Proposition 2: Consider a parallel structure as in Figure 1 with the number of sensors $N \geq 2$, the number of messages $D = 2$, and the number of hypotheses $M = 2$. When the observations of the sensors are conditionally dependent, there exists a fusion rule γ_0 in which the minimum average probability of error P_e given in (9) cannot be expressed solely as a function of $q(\gamma_1, \ldots, \gamma_N)$ (given in (5)).

Proof: We first prove this proposition for the 2-sensor case and then use induction to extend the result to $N > 2$. As before, let d_1 and d_2 denote the messages that sensor 1 and sensor 2 send to the fusion center. For notational simplicity, let $P_i(l_1, l_2)$ denote $P(d_1 = l_1, d_2 = l_2|H_i)$ where $l_1, l_2 \in \{0, 1\}$. We have the following linear system of equations with $P_i(0,0), P_i(0,1), P_i(1,0),$ and $P_i(1,1)$ as the unknowns.

$$\begin{aligned} P_i(0,0) + P_i(0,1) &= P_i(l_1 = 0) \\ P_i(1,0) + P_i(1,1) &= P_i(l_1 = 1) = 1 - P_i(l_1 = 0) \\ P_i(0,0) + P_i(1,0) &= P_i(l_2 = 0) \\ P_i(0,1) + P_i(1,1) &= P_i(l_2 = 1) = 1 - P_i(l_2 = 0) \end{aligned}$$

Note that the matrix of coefficients is singular. Solving this system, we have that

$$\begin{aligned} P_i(0,0) &= \alpha_i \\ P_i(0,1) &= P_i(l_1 = 0) - \alpha_i \\ P_i(1,0) &= P_i(l_2 = 0) - \alpha_i \\ P_i(1,1) &= 1 - P_i(l_1 = 0) - P_i(l_2 = 0) + \alpha_i \end{aligned}$$

where $\alpha_i, i = 0, 1$, corresponding to H_0, H_1 are real numbers in $(0, 1)$. Now we rewrite (9) for a fixed fusion rule γ_0:

$$P_e = \pi_0 \sum_{(d_1,d_2)\in R_1} P_0(d_1, d_2) + \pi_1 \sum_{(d_1,d_2)\in R_0} P_1(d_1, d_2) \quad (13)$$

where R_0 and R_1 are two partitions of the set of all possible values of (d_1, d_2) in which the fusion center decides hypothesis H_0 or hypothesis H_1 is true, respectively. Now suppose that the fusion center uses the following fusion rule: It picks 1 if $(d_1, d_2) = (1, 1)$ and picks 0 for the remaining three cases. After some manipulation, expression (13) becomes

$$\begin{aligned} P_e = \; & \pi_0(1 - P_0(d_1 = 0) - P_0(d_2 = 0) + \alpha_0) \\ & + \pi_1 (P_1(d_1 = 0) + P_1(d_2 = 0) - \alpha_1) \quad (14) \end{aligned}$$

From Proposition 1, α_0 is not uniquely determined given α_1 and vice versa. Thus P_e in (13) cannot be expressed solely as a function of $q(\gamma_1, \gamma_2)$.

Now we prove the proposition for $N > 2$ by induction on N. Suppose that there exists a fusion rule $\gamma_0^{(N)}$ that results in $P_e^{(N)}$ that cannot be expressed solely as a function of $q(\gamma_1,\ldots,\gamma_N)$; we will then show that there exists a fusion rule $\gamma_0^{(N+1)}$ that yields $P_e^{(N+1)}$ that cannot be expressed solely as a function of $q(\gamma_1,\ldots,\gamma_{N+1})$. Let $\tilde{R}_0^{(N)}$ and $\tilde{R}_1^{(N)}$ be the decision regions (for H_0 and H_1, respectively) at the fusion center when there are N sensors. Let $\tilde{R}_0^{(N+1)}$ and $\tilde{R}_1^{(N+1)}$ be those of the $(N+1)$-sensor case. Without loss of generality, we assume that the observation of sensor $(N+1)$ is independent of those of the first N sensors. Rewriting (9) for the N-sensor problem, we have that:

$$\begin{aligned} P_e^{(N)} &= \pi_0 \sum_{(l_1,\ldots,l_N)\in\tilde{R}_1^{(N)}} P_0(l_1,\ldots,l_N) \\ &\quad + \pi_1 \sum_{(l_1,\ldots,l_N)\in\tilde{R}_0^{(N)}} P_1(l_1,\ldots,l_N) \end{aligned}$$

Now we construct $\tilde{R}_0^{(N+1)}$ and $\tilde{R}_1^{(N+1)}$ based on $\tilde{R}_0^{(N)}$ and $\tilde{R}_1^{(N)}$ as follows. $\tilde{R}_0^{(N+1)}$ consists of combinations of the forms $(l_1,\ldots,l_N,0)$ and $(l_1,\ldots,l_N,1)$ where $(l_1,\ldots,l_N)\in\tilde{R}_0^{(N)}$; $\tilde{R}_1^{(N+1)}$ consists of combinations of the forms $(l_1,\ldots,l_N,0)$ and $(l_1,\ldots,l_N,1)$ where $(l_1,\ldots,l_N)\in\tilde{R}_1^{(N)}$. Note that, for $i=0,1$,

$$P_i(l_1,\ldots,l_N,0)+P_i(l_1,\ldots,l_N,1)=P_i(l_1,\ldots,l_N)\,.$$

Thus, P_e for the $(N+1)$-sensor case can be written as

$$\begin{aligned} P_e^{(N+1)} &= \pi_0 \sum_{(l_1,\ldots,l_N,l_{N+1})\in\tilde{R}_1^{(N+1)}} P_0(l_1,\ldots,l_N,l_{N+1}) \\ &\quad + \pi_1 \sum_{(l_1,\ldots,l_N,l_{N+1})\in\tilde{R}_0^{(N+1)}} P_1(l_1,\ldots,l_{N+1}) \\ &= \pi_0 \sum_{(l_1,\ldots,l_N)\in\tilde{R}_1^{(N)}} P_0(l_1,\ldots,l_N) \\ &\quad + \pi_1 \sum_{(l_1,\ldots,l_N)\in\tilde{R}_0^{(N)}} P_1(l_1,\ldots,l_N) = P_e^{(N)} \end{aligned}$$

But $P_e^{(N)}$ cannot be expressed solely as a function of $q(\gamma_1,\ldots,\gamma_N)$ and $q(\gamma_{N+1})$ due to the induction hypothesis and the independence assumption of sensor $(N+1)$'s observation. Thus $P_e^{(N+1)}$ cannot be expressed solely as a function of $q(\gamma_1,\ldots,\gamma_{N+1})$. ■

Thus, for the case of conditionally dependent observations, instead of using conditional marginal distributions, we relate the Bayesian probability of error to the joint distribution of the decisions of the sensors. In what follows, we use γ to collectively denote $(\gamma_1,\gamma_2,\ldots,\gamma_N)$ and Γ to denote the Cartesian product of $\Gamma_1,\Gamma_2,\ldots,\Gamma_N$, where Γ_j is the set of all deterministic decision rules (quantizers) of sensor j, $j=1,\ldots,N$. Also, we define

$$s_{d_1,\ldots,d_N}(\gamma|H_i)=Pr(\gamma_1=d_1,\ldots,\gamma_N=d_N|H_i) \quad (15)$$

Then, the D^N-tuple $s(\gamma|H_i)$ is defined as:

$$\begin{aligned} s(\gamma|H_i) = (&s_{0,0,\ldots,0}(\gamma|H_i), s_{0,0,\ldots,1}(\gamma|H_i),\ldots, \\ &s_{D-1,D-1,\ldots,D-1}(\gamma|H_i)) \end{aligned} \quad (16)$$

Finally, we define the $M\times D^N$-tuple $s(\gamma)$:

$$s(\gamma)=(s(\gamma|H_0),s(\gamma|H_1),\ldots,s(\gamma|H_{M-1})) \quad (17)$$

From (9), it can be seen that P_e is a continuous function on $s(\gamma)$ for a fixed fusion rule. We now prove that the set $S=\{s(\gamma):\gamma_1\in\Gamma_1,\ldots,\gamma_N\in\Gamma_N\}$ is compact, and therefore there exists an optimal solution for a fixed fusion rule. As the number of fusion rules is finite, we then can conclude that there exists an optimal solution for the whole system for each class of decision rules at the sensors.

Theorem 1: The set S given by

$$S=\{s(\gamma):\gamma_1\in\Gamma_1,\gamma_2\in\Gamma_2,\ldots,\gamma_N\in\Gamma_N\} \quad (18)$$

is compact.

Proof: To prove this theorem, we follow the same line of argument as in the proof of compactness of the set of conditional distributions for the one sensor case by Tsitsiklis [9]. Let $\mathcal{P}=(P_0+\ldots+P_{M-1})/M$, where $P_0,\ldots,P_{M-1}$ are the conditional distributions of the observations given $H_0,\ldots,H_{M-1}$, respectively. We use G to denote the set of all measurable functions from the observation space, $\mathcal{Y}=\mathcal{Y}_1\times\mathcal{Y}_2\times\ldots\times\mathcal{Y}_N$, into $\{0,1\}$. Let $G^{(D^N)}$ denote the Cartesian product of D^N replicas of G. Let

$$F = \left\{ (f_{00\ldots0},\ldots,f_{(D-1)(D-1)\ldots(D-1)})\in G^{(D^N)} \;\middle|\; \mathcal{P}\left(\sum_{d_1,\ldots,d_N=0}^{D-1} f_{d_1,\ldots,d_N}(Y)=1\right)=1 \right\} \quad (19)$$

For any $\gamma\in\Gamma$ and $d_1,\ldots,d_N\in\{0,\ldots,D-1\}$, we define $f_{d_1,\ldots,d_N}$ such that $f_{d_1,\ldots,d_N}(y)=1$ if and only if $\gamma(y)=(d_1,\ldots,d_N)$, and $f_{d_1,\ldots,d_N}(y)=0$ otherwise. Then, $f_{d_1,\ldots,d_N}$ will be the indicator function of the set $\gamma^{-1}(d_1,\ldots,d_N)$. It can be seen that $(f_{00\ldots0},\ldots,f_{(D-1)(D-1)\ldots(D-1)})\in F$. Also, we have

$$\begin{aligned} s_{d_1,\ldots,d_N}(\gamma|H_i) &= Pr(\gamma(y)=(d_1,\ldots,d_N)|H_i) \\ &= \int f_{d_1,\ldots,d_N}(y)dP_i(y). \end{aligned} \quad (20)$$

Conversely, for any $f=(f_{00\ldots0},\ldots,f_{(D-1)(D-1)\ldots(D-1)})\in F$, define $\gamma\in\Gamma$ as follows.

- If $\sum_{d_1,\ldots,d_N=0}^{D-1} f_{d_1,\ldots,d_N}(y)=1$, then $\gamma(y)=(d_1,\ldots,d_N)$ such that $f_{d_1,\ldots,d_N}(y)=1$.
- If $\sum_{d_1,\ldots,d_N=1}^{D} f_{d_1,\ldots,d_N}(y)\neq 1$, then $\gamma(y)=(1,1,\ldots,1)$.

As $\mathcal{P}\left(\sum_{d_1,\ldots,d_N=1}^{D} f_{d_1,\ldots,d_N}(Y)\neq 1\right)=0$, (20) still holds. Now we define a mapping $h:F\to\Re^{MD^N}$ such that

$$h_{i,d_1,\ldots,d_N}(f)=\int f_{d_1,\ldots,d_N}dP_i(y) \quad (21)$$

It can be seen that $S=h(F)$. If we can find a topology on G in which F is compact and h is continuous, S will be a compact set.

Let $L_1(\mathcal{Y};\mathcal{P})$ denote the set of all measurable functions $f:\mathcal{Y}\to\mathcal{R}$ that satisfy $\int|f(y)|d\mathcal{P}(y)<\infty$, $L_\infty(\mathcal{Y};\mathcal{P})$ denote the set of all measurable functions $f:\mathcal{Y}\to\mathcal{R}$

such that f is bounded after removing the set $\mathcal{Y}_z \subset \mathcal{Y}$ that has $\mathcal{P}(\mathcal{Y}_z) = 0$. Then G is a subset of $L_\infty(\mathcal{Y};\mathcal{P})$. It is known that $L_\infty(\mathcal{Y};\mathcal{P})$ is the dual of $L_1(\mathcal{Y};\mathcal{P})$ [14]. Consider the weak* topology on $L_\infty(\mathcal{Y};\mathcal{P})$, which is the weakest topology where the mapping

$$f \rightarrow \int f(y)g(y)d\mathcal{P}(y) \tag{22}$$

is continuous for every $g \in L_1(\mathcal{Y};\mathcal{P})$. Using Alaoglu's theorem [14], we have that the unit ball in $L_\infty(\mathcal{Y};\mathcal{P})$ is weak*-compact. Thus G is compact. Then $G^{(D^N)}$, which is a Cartesian product of D^N compact sets, is also compact. Now, from (19), every point $\left(f_{00\ldots 0}, \ldots, f_{(D-1)(D-1)\ldots(D-1)}\right) \in F$ satisfies

$$\int_A \sum_{d_1,\ldots,d_N=0}^{D-1} f_{d_1,\ldots,d_N}(y)d\mathcal{P}(y) = \mathcal{P}(A), \tag{23}$$

where A is any measurable subset of $\mathcal{Y}$. If we let X_A denote the indicator function of A, it follows that

$$\int \sum_{d_1,\ldots,d_N=0}^{D-1} f_{d_1,\ldots,d_N}(y)X_A(y)d\mathcal{P}(y) = \mathcal{P}(A). \tag{24}$$

As $X_A \in L_1(\mathcal{Y};\mathcal{P})$ and the mapping in (22) is continuous for every $g \in L_1(\mathcal{Y};\mathcal{P})$, we have that the map $f \rightarrow \mathcal{P}(A)$ is also continuous. Furthermore, F is a subset of the compact set $G^{(D^N)}$, and thus F is also compact.

Let $g_i, i = 0, \ldots, M-1$ denote the Radon-Nikodym derivative of P_i with respect to $\mathcal{P}$, $g_i(y) = \frac{dP_i(y)}{d\mathcal{P}(y)}$. Then we have $g_i \in L_1(\mathcal{Y};\mathcal{P})$ [9]. Also, we have that

$$\int f_{d_1,\ldots,d_N}(y)dP_i(y) = \int f_{d_1,\ldots,d_N}(y)g_i(y)d\mathcal{P}(y), \quad \forall i, d_1, \ldots, d_N. \tag{25}$$

From (22), (25) and the fact that $g_i \in L_1(\mathcal{Y};\mathcal{P})$, it follows that the mapping $f \rightarrow \int f_{d_1,\ldots,d_N}(y)dP_i(y)$ is continuous. Therefore the mapping h given in (21) is continuous. As $S = h(F)$, we finally have that S is compact. ■

Theorem 2: There exists an optimal solution for the general rules at the sensors, and there also exists an optimal solution for the special case where the sensors are restricted to the threshold rules on likelihood ratios.

Proof: For each fixed fusion rule γ_0 at the fusion center, the probability of error P_e given in (9) is a continuous function on the compact set S. Thus, by Weierstrass theorem [14], there exists an optimal solution that minimizes P_e for each γ_0. Furthermore, there is a finite number of fusion rules γ_0 at the fusion center (in particular, this is the number of ways to partition the set $\{d_1, d_2, \ldots, d_N\}$ into two subsets, which is 2^N). Therefore, there exists an optimal solution over all the fusion rules at the fusion center. Note that the use of the general rule or the threshold rule will result in different fusion rules, but will not affect the reasoning in this proof. The optimal solutions in each case, however, will be different in general. More specifically, the set of all the decision rules (of the sensors) based on the threshold rule will be a subset of the set of all decision rules (of the sensors), thus the optimal solution in the former case will be worse than that of the latter in general. ■

IV. A SPECIAL CASE WITH BIVARIATE NORMAL DISTRIBUTIONS AND SIMULATION RESULTS

In this section, we consider a special case with $M = 2$, $N = 2$, $D = 2$, and the joint distribution given each hypothesis is bivariate normal. Particularly, let the joint distribution of the observations given each hypothesis be $f_0(y_1, y_2)$ (given H_0), which is a bivariate normal density with means $\mu_1 = \mu_2 = -1$, variances $\sigma_1^2 = \sigma_2^2 = 1$, the correlation coefficient $\rho = 0.6$, and $f_1(y_1, y_2)$ (given H_1), which is also a bivariate normal density, with $\mu_1 = \mu_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho = 0.6$. These two distributions are plotted in Figure 3. Here, $\mathcal{Y}_j \equiv \mathcal{R}$, for $j = 1, 2$. Note that even when the observations are *i.i.d.*, restricting the sensors to the same decision rules may lead to a suboptimal solution [4]. Thus, we do not assume that the decision rules of the two sensors are the same for the simulations in this section. In what follows, we derive some properties of the minimum P_e and present some numerical results for both threshold rules and general rules at the sensors.

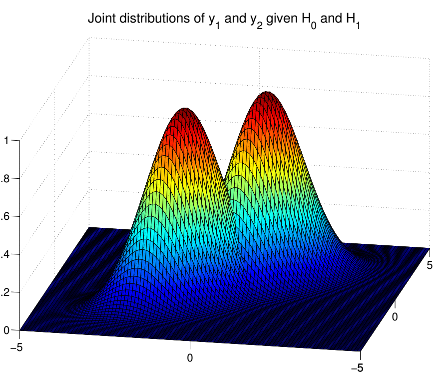

Fig. 3. Joint distributions of Y_1 and Y_2 given H_0 and H_1

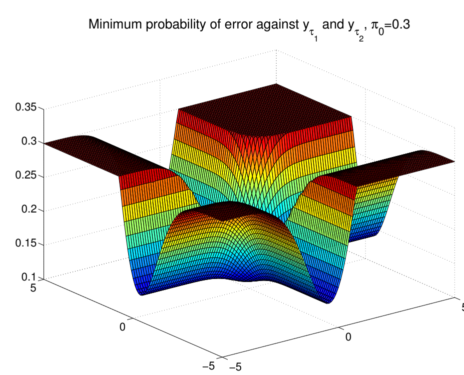

Fig. 4. Minimum probabilities of error versus y_{τ_1} and y_{τ_2} with $\pi_0 = 0.3$

A. Using Threshold Rules at the Sensors

At each sensor, the marginal distribution of the observation is Gaussian with variance $\sigma^2 = 1$ and mean -1 under H_0

and mean 1 under H_1. The (marginal) likelihood ratios are monotonically increasing in y_1 and y_2, respectively, thus a threshold rule for the likelihood ratios becomes a threshold rule for y_1 and y_2.

$$\gamma_j(y_j) = \begin{cases} 1 & \text{if } y_j \geq y_{\tau_j} = \sigma^2 \ln \tau_j / 2 \\ 0 & \text{otherwise,} \end{cases} \quad (26)$$

The conditional joint distributions of sensor messages are given by (10), where

$$\begin{aligned} N &= 2, R_0^{(1)} = (-\infty, y_{\tau_1}), R_1^{(1)} = [y_{\tau_1}, \infty), \\ R_0^{(2)} &= (-\infty, y_{\tau_2}), R_1^{(2)} = [y_{\tau_2}, \infty). \end{aligned} \quad (27)$$

As $\int_{-5}^{5} \int_{-5}^{5} f_i(y_1, y_2) dy_1 dy_2 \approx 0.9999$ for $i = 0, 1$, it suffices to let y_{τ_j} vary within $[-5, 5]$. We then use equally spaced values of y_{τ_j} as threshold candidates. The minimum values of P_e using threshold rules with $\pi_0 = 0.3$ are plotted in Figure 4. From the simulation results, it can be observed that:

$$P_e \leq \min\{\pi_0, \pi_1\}, \lim_{y_{\tau_1}, y_{\tau_2} \to \pm\infty} P_e = \min\{\pi_0, \pi_1\}. \quad (28)$$

We state below a generalization of these observations.

Proposition 3: Consider a parallel structure as in Figure 1 with the number of sensors $N = 2$, the number of messages $D = 2$, and the number of hypotheses $M = 2$. Let $f_0(y_1, y_2)$ and $f_1(y_1, y_2)$ be the joint probability density functions of the sensor observations given H_0 and H_1, respectively, where $f_i(y_1, y_2)$, $i = 0, 1$, are continuous on $\mathcal{R}^2$ and nonzero for $-\infty < y_1, y_2 < \infty$. Assume further that the decision regions of each sensor are of the form $R_0^{(j)} = (-\infty, y_{\tau_j})$ and $R_1^{(j)} = [y_{\tau_j}, +\infty)$, $y_{\tau_j} \in (-\infty, +\infty)$, where $j = 0, 1$ (which are threshold rules on the observation values). Then we have (28) where P_e is given in (9).

Proof: The proof of this proposition can be found in the full version of this paper [15]. ■

B. Using General Rules at the Sensors

The observation space of each sensor ($\mathcal{Y}_j$) is partitioned into two decision regions, $R_0^{(j)}$ and $R_1^{(j)}$. Particularly, we first divide $\mathcal{Y}_j$ into I_j intervals. Then there will be 2^{I_j} different ways to partition these intervals into $R_0^{(j)}$ and $R_1^{(j)}$. To go through all of these possibilities, we use an I_j-bit counter where the n^{th} bit, $n = 0, \ldots, I_j - 1$, indicates which region the corresponding interval resides in. The conditional joint distributions of sensor messages are given by (10), where $N = 2$. In the simulations we have carried out (whose results can be found in [15]), the general rule leads to the same optimal solutions as the threshold rule.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have shown that the minimum Bayesian probability of error P_e in a parallel configuration cannot be expressed as a function of the conditional marginal distributions of the messages from the sensors. We have then characterized this probability of error based on the set of conditional joint distributions. We have proved that this set is compact and therefore there exist optimal solutions that minimize P_e for both the general decision rules and the threshold rules at the sensors. We have also carried out simulations for a special case where the joint distributions of the sensor observations are bivariate normal. Within the values of the parameters simulated, the results have shown that the threshold rules at the sensors achieve the optimal P_e of the general rules.

As mentioned earlier, in the applications of decentralized detection such as sensor networks and network security, sensor observations may be correlated given each hypothesis. Characterizing the sensors based on the conditional joint distributions will open up a new avenue for solving decentralized detection problems.

VI. ACKNOWLEDGMENTS

We would like to thank Deutsche Telekom Laboratories and the Vietnam Education Foundation for their support. We are also grateful to three anonymous reviewers for their valuable comments.

REFERENCES

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

[2] R. R. Tenney and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES–17, pp. 501-510, 1981.

[3] F. A. Sadjadi, "Hypothesis testing in a distributed environment," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-22, No. 2, March 1986, pp. 134-137.

[4] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Signal Processing*, JAI Press, 1993.

[5] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transaction on Automatic Control*, Vol. AC–30, No. 5, May 1985, pp. 440–446.

[6] I. Y. Hoballah and P. K. Varshney,"Distributed Bayesian signal detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 995–1000, Sept. 1989.

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

[8] A. Kashyap, T. Başar, and R. Srikant, "Asymptotically Optimal Quantization for Detection in Power Constrained Decentralized Networks," *Proceedings of the 2006 American Control Conference*, Minnesota, USA, June 14-16, 2006.

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

[10] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5099–5106, Nov. 2006.

[11] K. C. Nguyen, T. Alpcan, and T. Başar, "A Decentralized Bayesian Attack Detection Algorithm for Network Security," *Proc. of the 23rd International Information Security Conference (SEC 2008)*, Milan, Italy, Sep. 2008.

[12] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad, and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Process.*, vol. 48, pp. 3266-3279, Dec. 2000.

[13] J. Unnikrishnan and V. V. Veeravalli, "Decentralized Detection with Correlated Observations," in *Proc. of Asilomar Conference on Signals, Systems, and Computers*, Nov. 2007.

[14] D. G. Luenberger, *Optimization by Vector Space Methods*, New York: John Wiley & Sons, 1969.

[15] K. C. Nguyen, T. Alpcan, and T. Başar, "Distributed Hypothesis Testing with a Fusion Center: The Conditionally Dependent Case," *Technical Report*, UIUC, Aug. 2008. Available at http://decision.csl.uiuc.edu/˜knguyen4/research/research.html