An Efficient Approximation Scheme for Distributed Multiple Hypothesis Testing [1]

Une stratégie efficace d'approximation pour un test d'hypothèses multiples distribuées

par Jason D. PAPASTAVROU
School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA.

Abstract

In the general distributed detection problem a set of decision makers (DMs) receive observations of the environment and transmit finite-valued messages to other DMs according to prespecified communication protocols. A designated primary DM makes the final decision on one of the alternative hypotheses. All DMs make decisions so as to optimize a measure of organizational performance. Since the "quest for optimality" in problems in this framework is associated with great computational and inherent complexity, simple approximate solutions which take into consideration the specific characteristics of the problem should be employed. This approach is demonstrated by addressing the issues involved with reducing a complex M-ary hypothesis testing problem into a sequence of simpler binary hypothesis testing subproblems. An approximate decision scheme is derived that is computationally easy to implement and performs very well by exploiting the structure of the alternative hypotheses of each particular problem.

Key words : Distributed Multiple Hypothesis Testing, Approximation Scheme

Résumé

Dans le problème général de la deetection distribuée, un ensemble de preneurs de décision reçoit des observations de l'environnement et transmet des messages prenant des valeurs finies à d'autres preneurs de décision selon des protocoles de communication préétablis. Un preneur de décision choisi comme primaire, prend la décision finale sur l'une des hypothèses possibles. Tous les preneurs de décision décident de manière à optimiser une mesure de performance d'organisation. Puisque la «quête d'optimalité» dans ce type de problèmes s'accompagne d'une grande complexité et d'une lourde charge de calcul, des solutions approximatives simples qui tiennent compte des caractéristiques spécifiques du problème devraient être employées. Nous présentons cette approche en convertissant les questions impliquées dans la réduction d'un problème complexe de test d'hypothèses M-aires en une suite de sous-problèmes de test d'hypothèses binaires. Nous derivons une stratégie de decision approximative facile à implanter numériquement et qui est très efficace car elle utilise la structure des hypothèses possibles dans chaque problème particulier.

Mots clés : *Test d'hypothèses multiples distribuées, Stratégie d'approximation*

1. Introduction and Motivation

The problem of distributed decision making in a hypothesis testing environment has attracted considerable interest during the past decade. This framework was selected because it combines two desirable attributes; the mathematical problems are easy to describe so that researchers from diverse disciplines can understand the models and their conclusions; also, the problems have trivial centralized counterparts, so that all the difficulties arise because of the decentralization of the decision making process. On the other hand, these problems are also known to become computationally intractable (NP-hard) even for a small number of decision makers (DMs) and a small number of communication messages [25]. Thus, in order to overcome the limitations caused by the combinatorial complexity, it would be desirable to develop computationally efficient approximate procedures, which hopefully will not cause significant deterioration in the performance of the organization.

We examine problems of cooperative organizations which consist of a number of DMs and perform M-ary hypothesis testing. Even though we attempt to keep the combinatorial complexity under control, considerable difficulties arise from the intrinsic complexity of the distributed problems. We present alternative decision schemes for these organizations, analyze them in a quantitative manner and compare their performance and computational requirements to the performance and computational requirements of the optimal decision rules. We investigate whether some "common sense" and "intuitively appealing" procedures, which are easy to

1. This research was supported by the National Science Foundation under grant DDM-9309579.

implement, perform satisfactorily in this framework. (Similar approaches that analyzed other aspects of the decision making process were followed in [12] and [13].) The objective is to analyze the well-known trade-off : optimality versus computational efficiency. We argue that the proposed decision scheme performs very well compared to the optimal, and therefore should be preferred due to the tremendous reduction in computational complexity.

The Bayesian decentralized detection problem was first considered in [24], where the optimality of constant threshold strategies was established; this was formalized and generalized in [27]. Several generalizations of the basic detection model have appeared in [5], [22], [1], [2], [8] and [6]. The parallel architecture with identical sensors has been analyzed in [8], [18] and in [19]. An explicit asymptotic solution for a special case of the parallel architecture performing M-ary hypothesis testing appeared in [16]. Multiple hypotheses were also considered in [17] and [23]. Asymptotic results were established in [25] for the parallel architecture and in [12] for the tandem (serial) architecture. The Neyman-Pearson formulation of similar problems is considered in [7], [19], [20], [7] and [29]. The optimality of a monotone threshold strategy for the case of independently randomized decision rules was established in [30]. The effects of randomized decision rules were considered in [14]. In [33], different team architectures are compared for the Neyman-Pearson formulation. Different team architectures, for the Bayesian case, are also compared in [19] (numerically) and very extensively in [13] (analytically). In [11] and [32], the effects of communication costs were considered. In [15] and [31], hypothesis testing models are used to analyze sensor decision integration. Two thorough overviews of the field appeared in [28] and [3].

In section 2, the distributed hypothesis testing framework is defined and the M-ary distributed hypothesis testing problem is formally presented together with the optimal solution. The team has to select one of M alternative hypotheses, taking into account different costs for hypothesis misclassification. The team consists of N DMs; one of them is referred to as the primary DM and is responsible for the final team decision, and the rest are known as the consultant DMs (Figure 1). Each consultant DM receives his/her observation and then computes and communicates a U-ary message to the primary DM. Upon receipt of his/her own observation and the messages from the consultant DMs, the primary DM makes the final team decision declaring one of the M hypotheses to be true. In general, in order to determine the optimal team configuration (i.e., the optimal primary DM), the problem needs to be solved N times; each time with a different primary DM.

In section 3, an approximate decision scheme is presented which has a dual objective. The first objective is to make an educated selection, among the DMs of the team, for the DM that will be designated as the primary DM. The second objective is to decrease the computational complexity of the original problem by reducing the complex M-ary hypothesis testing problem into a series of binary hypothesis testing problems. This is achieved through a corrected adaptation of the scheme in [4] for object recognition with a single sensor. Furthermore, several suggestions are made for improving the approximate decision scheme by taking maximum advantage of the particular structure of the problem; thus the philosophy and benefits of the proposed problem dependent analysis are demonstrated. In section 4, we analytically discuss two numerical examples that demonstrate the efficiency of the approximate decision scheme and in section 5 we present our conclusions.

2. The M-ary Distributed Hypothesis Testing Problem

It is well-known that the combinatorial complexity of the distributed hypothesis testing problems almost always explodes as the number of the team DMs, or of the hypotheses, or of the messages increases. On the other hand, simple, approximate and usually intuitive solutions have been shown to perform relatively well. Moreover, in [13] it was shown that the optimal team architecture in general depends on the characteristics of the environment. Therefore, it becomes apparent that it is worthwhile for the designer of the team to : (i) sacrifice some of the team performance, in order to keep the complexity under control and, (ii) take into consideration the specific characteristics of the environment in which the team operates, in order to improve the team performance. To demonstrate this approach, a complex multiple hypothesis testing problem with a team consisting of several DMs is presented.

PROBLEM. Consider a team which consists of N DMs and performs M-ary hypothesis testing. Each hypothesis H_m occurs with a known prior probability $p_m = P(H_m)$, for $m = 0, 1, ..., M-1$. DM $n (n = 1, 2, ..., N)$ receives an observation y_n with a corresponding probability density function $P(y_n|H_m)$ given hypothesis H_m. The observations are conditional independent given the true hypothesis. One DM is designated as the primary DM and the others are referred to as the consultant DMs (Figure 1). Each consultant DM n transmits to the primary DM an U-ary message $u_n \in \{0, 1, ..., U-1\}$ based on his/her observation y_n. The primary DM considers his/her own observation and the communications from the consultant DMs and makes the final team decision $u_1 \in \{0, 1, ..., M-1\}$, declaring one of the hypotheses to be true. There exists a cost $J(u_1, H)$ associated with the team deciding u_1 when H is the true hypothesis[2]. The objective is to determine the decision rules for the DMs that minimize the team probability of error taking into account the different costs for hypothesis misclassification.

2. Throughout this discussion we assume that the cost function is such that it is more costly for the team to err than to be correct (i.e., $J(k, H_m) > J(m, H_m)$, for all $k, m = 0, 1, ..., M-1$ with $k \neq m$). This logical assumption is made in order to express the optimal decision rules in the convenient likelihood ratio form.

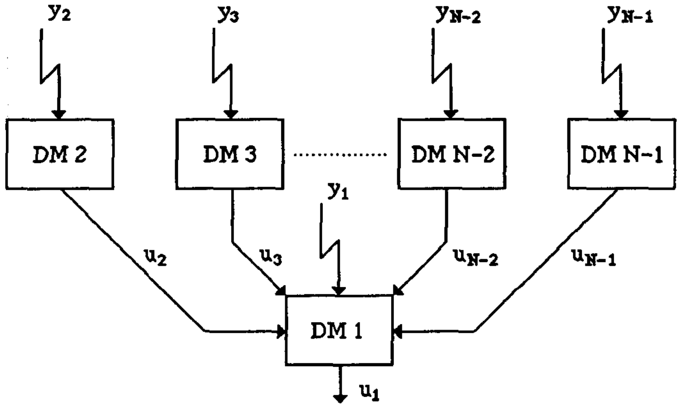

Figure 1. – The Team of the M-ary Distributed Hypothesis Testing Problem.

Given the communication vector $\boldsymbol{u} = (u_2, u_3, ..., u_N) \in \{0, 1, ..., U-1\}^{N-1}$ transmitted from the consultant DMs to the primary DM, the optimal decision rule of the primary DM is given by the following likelihood ratio tests with constant thresholds :

$$\sum_{m=1}^{M-1} \Lambda_m(y_1)p_m[J(i,H_m)-J(j,H_m)]\sum_{n=2}^{N}\frac{P(u_n\mid Hm)}{P(u_n\mid H_0)} \underset{u_1\neq j}{\overset{u_1\neq i}{\gtrless}} p_0[J(j,H_0)-J(i,H_0)] \quad (1)$$

for : $i = 1, 2, ..., M-2$, and $j = m+1, m+2, ..., M-1$, where :

$$\Lambda_m(y_n) \equiv \frac{P(y_n \mid H_m)}{P(y_n \mid H_0)}\,; \qquad m = 1, 2, ..., M-1;\; n = 1, 2, ..., N. \quad (2)$$

The optimal decision rules of the consultant DMs are given by :

$$\sum_{m=1}^{M=1} \Lambda_m(y_n)p_m \sum_{u_1=0}^{M-1}[P(u_1 \mid u_n = i, H_m) - P(u_1 \mid u_n = j, H_m)]J(u_1, H_m) \underset{u_n\neq j}{\overset{u_n\neq i}{\gtrless}}$$

$$\underset{u_n\neq j}{\overset{u_n\neq i}{\gtrless}} p_0 \sum_{u_1=0}^{M-1}[P(u_1 \mid u_n = j, H_0) - P(u_1 \mid u_n = i, H_0)]J(u_1, H_0) \quad (3)$$

for : $i = 0, 1, ..., U-2; j = i+1, i+2, U-1; n = 2, 3, ..., N.$

Suppose that the optimal team configuration (that is, the optimal placement of the DMs within a given architecture) is to be determined for some given prior probabilities and some given costs. The M-ary distributed hypothesis testing problem needs to be solved N different times each time with a different primary DM to obtain the optimal team performance. Recalling the results of [13] for a team with a single consultant DM performing binary hypothesis testing, one concludes that the optimal configuration for the team can only be determined in this manner because it depends both on the prior probabilities and on the associated costs.

Remark 1. The U-ary messages from the consultant DMs to the primary DM are denoted as $0, 1, ..., U-1$ for the sake of simplicity. For that matter they can be denoted with any U distinct symbols or names. Note that consultant DM transmitting a 0 message does not necessarily indicate that H_0 is the correct hypothesis. The important fact is that each consultant DM has a U-ary message in his/her disposition and should try, together with the primary DM, to make optimal use of them so that the team expected cost be minimized.

Remark 2. In the sequel, we set $U = M$; that is, the number of messages that each consultant DM can transmit to the primary DM is equal to the number of the hypotheses. In [17], it was suggested that this a sufficient number of messages for the team to perform well relative to its centralized counterpart.

A Gauss-Seidel algorithm for determining the optimal decision rules was presented in [23]. But, the optimality conditions of Eqs. (1) and (3) are just *necessary* conditions. If there exists a suboptimal set of decision rules for the team which satisfies these conditions, then the algorithm could converge to a suboptimal cost. Even if we assume that this does not occur for some 'well behaved' probability density functions, implementation of the algorithm for six or more hypotheses is highly non-trivial even if the consultant DM can only transmit binary messages; it is hard to determine the two five-dimensional decision regions and calculate the corresponding probabilities. In fact, we do not have any feeling for what these regions would look like. It should be an interesting problem for future research to implement an algorithm which solves a (Gaussian) six hypothesis testing problem for a team which consists of six DMs and binary communications. Since the thresholds on the likelihood ratio hyperplane can be translated to thresholds in the observation axis, this should provide some insight on how information should be summarized and fused.

In conclusion, the determination of the optimal decision rules and their associated decision regions is highly non-trivial, even for a small number of alternative hypotheses and communication messages; this suggests that it should be worthwhile to abandon optimality for an approximate solution which is easily obtainable. This is in accordance to a very interesting and widely accepted theory of Simon; the theory of *satisficing* [21].

This theory claims that decision makers do not search for the optimal decision, but rather for a decision which is "good enough." Simon sees decision makers as agents of *bounded or limited rationality.* The bounds of rationality depend on several factors like the expertise of the decision makers involved, the importance of the intellectual task and the availability of relevant tools. Decision makers will be satisfied with a decision which yields even a small improvement compared to the present and will try to avoid uncertainty. They will usually settle for a smaller expected pay-off if this implies a smaller variance. Decision makers are content to look at a drastically simplified model of the world and to examine just two alternatives, if these meet some minimal requirements : the same course and one new course. Therefore, this further suggests that it can be very beneficial to reduce the complex M-ary distributed hypothesis testing problem into a series of (simple) *binary* hypothesis testing problems. This is exactly what the approximate decision scheme of the following section is trying to accomplish.

3. The Approximate Decision Scheme

In this section an approximate decision scheme for the M-ary distributed hypothesis testing problem is presented and analyzed. As was explained in the previous section, the two main objectives of this decision scheme is (a) to significantly reduce the exploding computational complexity of the original problem, and (b) to maintain a near-optimal performance.

To achieve the first objective the M-ary distributed hypothesis testing problem is reduced into a series of binary hypothesis testing problems, where the hypotheses are not some of the original (single) hypotheses of the M-ary hypothesis testing problem, but rather *composite* hypotheses (each composite hypothesis consists of a set of the (single) hypotheses). For example, consider the problem of a radar operator who has to determine presence or absence (hypothesis H_0) of an airplane. If a plane is present the radar operator has also to decide whether it is an F-16 (H_1), a Mirage (H_2), an Airbus (H_3), or a Boeing (H_4). The radar operator may first consider the problem of the presence or absence of an airplane; the presence of the airplane corresponds to the composite hypothesis H_1, H_2, H_3 and H_4, and the absence of the target corresponds to H_0. If the operator decides that a plane is indeed present, then the operator may consider the binary problem "military airplame" (H_1 and H_2) versus "commercial airplame" (H_3 and H_4). Finally, depending on the operator's decision on the second binary composite hypothesis testing problem, the operator has to consider a final binary hypothesis testing problem (either H_1 versus H_2, or H_3 versus H_4).

The composite hypotheses are formally defined in the following section. But, in order to have a well-defined binary hypothesis testing problem the "prior" probabilities and the "misclassification" costs for these hypotheses need to be defined. For this reason, the *normalized prior probabilities*, that take into account both the original prior probabilities and the original misclassification costs of the M-ary hypothesis testing problem, are introduced now.

3.1. THE NORMALIZED PRIOR PROBABILITIES

(i). Elementary Hypotheses

The definition of the decision threshold for *binary* hypothesis testing is well documented in the literature (for example, see [13]):

$$\eta \equiv \frac{P(H_0)[J(1,H_0) - J(0,H_0)]}{P(H_1)[J(0,H_1) - J(1;H_1)]} \tag{4}$$

From Eq. (4), it is clear that the ratio of the additional costs incurred by the team when it makes the wrong decision influences the decision in the same way that the ratio of the prior probabilities does; for example, doubling the additional cost incurred by the team when it makes the wrong decision under hypothesis H is interpreted by the team, in the optimal decision rules, as doubling the relative frequency of occurrence of H. We try to extend this notion to the M-ary hypothesis case by introducing the *normalized prior probabilities.*

We try to summarize all of the information given by the prior probabilities and the misclassification costs into "relative frequencies" and thus define the (elementary) normalized prior probabilities as follows :

$$\pi_m \equiv \frac{p_m \sum_{i=0}^{M-1} \frac{p_u}{1-p_m}[J(u,H_m) - J(m,H_m)]}{\sum_{q=0}^{M-1} p_q \sum_{u=0}^{M-1} \frac{p_u}{1-p_q}[J(u,H_q) - J(q,H_q)]}; \quad \text{for } m = 0,1,...,M_1 \tag{5}$$

The restrictive assumption, implied in the above definition, is that given the true hypothesis H the relative frequencies of the different types of error are given by the prior probabilities.

Note that for the binary hypothesis testing (where given the true hypothesis *only one* type of error may occur), the normalized prior probabilities are "optimal" in the sense that their ratio is equal to the decision threshold η (i.e., $\eta = \pi_0/\pi_1$). Furthermore, in the case of the minimum error cost function (i.e., $J(u,H_m) = 1$ if $u \neq m$, and $J(u,H_m) = 0$ if $u = m$) the normalized prior probabilities reduce to the actual prior probabilities; since all errors are equally costly the relative frequencies of the hypotheses are not affected.

(ii). Composite Hypotheses

Consider $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$, two distinct subsets of $\{0,1,...,M-1\}$, and the binary hypothesis testing problem between the following two *composite* hypotheses, that is hypotheses which consist of a set of elementary hypotheses :

$$\boldsymbol{H}_{\boldsymbol{F}_0} \sum_{m\in\boldsymbol{F}_0} \frac{p_m}{\sum_{q\in\boldsymbol{F}_0} p_q} P(y \mid H_m) \quad \text{vs. } \boldsymbol{H}_{\boldsymbol{F}_1} : \sum_{m\in\boldsymbol{F}_1} \frac{p_m}{\sum_{q\in\boldsymbol{F}_1} p_q} P(y \mid H_m) \tag{6}$$

For this binary problem the hypotheses occur with prior probabilities :

$$p_{\boldsymbol{F}_i\boldsymbol{F}_{1-i}} = \frac{\sum_{m\in\boldsymbol{F}_i} p_m}{\sum_{q\in\boldsymbol{F}_0\cup\boldsymbol{F}_1} p_q}; \text{ for } i = 0,1. \tag{7}$$

The normalized prior probability of F_i for this binary problem is defined as :

$$\pi_{F_iF_{1-i}} = \frac{\sum\limits_{m\in F_i} p_m \sum\limits_{u\in F_{1-i}} \frac{p_u}{\sum\limits_{q\in F_{1-i}} p_q}[J(u,H_m) - J(m,H_m)]}{\sum\limits_{j=0}^{1}\sum\limits_{m\in F_j} p_m \sum\limits_{u\in F_{1-j}} \frac{p_u}{\sum\limits_{q\in F_{1-j}} p_q}[J(u,H_m) - J(m,H_m)]} \quad (8)$$

The two sets F_0 and F_1 of the hypotheses have to be mutually exclusive, but do not have to be collectively exhaustive (i.e., $F_0 \cap F_1 = \emptyset F_0 \cup F_1 \supseteq \{0, 1, ..., M-1\}$). Note that in the case of the minimum error cost function, the normalized prior probability of F_i of Eq. (8) reduces to the prior probability of F_i of Eq. (7). Moreover, the decision threshold for the composite binary hypothesis testing problem is given by :

$$\eta(F_0, F_1) = \frac{\sum\limits_{q\in F_0} p_q \sum\limits_{m\in F_0} p_m \sum\limits_{u\in F_1} p_u[J(u,H_m) - J(m,H_m)]}{\sum\limits_{q\in F_1} p_q \sum\limits_{m\in F_1} p_m \sum\limits_{u\in F_0} p_u[J(u,H_m) - J(m,H_m)]} \quad (9)$$

Hence just like in the usual binary case, the decision threshold can be broken down as the product of two ratios; the first being a ratio of the prior probabilities and the second being a ratio of the (weighted) additional costs.

3.2. THE BINARY DECISION TREE

Consider the following approximate decision scheme for a single DM to perform M-ary detection. The multiple hypothesis testing problem will be broken into $M-1$ binary hypothesis testing problems. For this an appropriate *binary decision tree*[3] is constructed (Figure 2). Consider any tree with M leaves (i.e., terminal nodes) having the following property : there exists a single non-terminal node that has exactly two edges emanating from it, and the rest of the non-terminal nodes have exactly three edges emanating from them. It is not hard to see inductively that such a tree contains exactly $M-1$ non-terminal nodes. The special non-terminal node with just the two edges is referred to as the *source node* and the other $M-1$ non-terminal nodes are referred to as the *decision* nodes; also every one of the M terminal nodes corresponds to one of the M hypotheses. The hypotheses can be assigned to the terminal nodes in any order.

3. A *tree* is a connected acyclic graph. A *binary decision tree* is defined as a directed tree which has the following three properties : one *source* node (i.e., a node with no arcs directed into it); there are exactly two directed arcs emanating from each non-terminal node; each teminal node is a destination (i.e., the node's single arc is directed into it). For a complete treatment of graph theory, the reader is refered to [PS82].

In the decision tree, there exists a unique (directed) path from the source node to each and every of the terminal nodes; in fact, starting from the source node any hypothesis can be declared (i.e., any terminal node can be reached) with a series of *binary* decisions. There is a total of $M-1$ binary decisions that can be made; one in each decision node of the tree. Consider any node t (except the source node) of the decision tree. The immediate predecessor of t on the path from the source node to t is called the *parent* node of t[4]. Similarly, consider any non-terminal node t; the two nodes t_0 and t_1 adjacent to t, which are not the parent of t, are called the *children* nodes of t. Finally, for any node t, the terminal nodes of the subtree with t as the source node is referred to as the *feasible* set of t and is denoted by $F(t)$. Note, that the feasible set of the source node consists of all the hypotheses H_0, H_1, ..., H_{M-1}, and that the feasible set of any terminal node consists of the one hypothesis that has been assigned to that terminal node.

Consider any non-terminal node t and denote by t_0 and t_1 its two children. Then, *the binary decision that is to be made at node t, is conditional on one of the hypotheses on the feasible set of t being true*, and requires the selection between two *composite* hypotheses : the feasible set of t_0 and the feasible set of t_1; that is :

$$H_{t_0} : \sum_{m\in F(t_0)} \frac{p_m}{\sum\limits_{q\in F(t_0)} p_q} P(y \mid H_m)$$
$$\text{vs. } H_{t_1} : \sum_{m\in F(t_1)} \frac{p_m}{\sum\limits_{q\in F(t_1)} p_q} P(y \mid H_m) \quad (10)$$

The prior probabilities for the hypotheses were defined in Eq. (7), as was the decision threshold in Eq. (9). Thus, at every decision node t, the primary DM will be employing a single likelihood ratio test by comparing the likelihood ratio of the probability density functions defined in Eq. (10) to the decision threshold of Eq. (9).

3.3. THE APPROXIMATE ALGORITHM

An approximate algorithm is now proposed for solving the M-ary distributed hypothesis testing problem of section 2. Thus the team architecture with a single primary DM, the consultant DMs and the binary communications is preserved so that the structure of the original problem be preserved. In order to compensate for the extra error incurred by the team, the algorithm needs to be computationally simple and easy to implement.

Suppose for the moment that a specific configuration for the team is given (i.e., the primary DM has been specified); this assumption will be relaxed in the sequel. Moreover, suppose that there exist exactly $M-1$ consultant DMs. This is not a really restrictive assumption. If the actual number of consultant DMs is less than $M-1$, some "dumb" (totally worthless) DMs can be added to the

4. The source node does not have a parent node.

team; if more than $M-1$ consultant DMs exist, the primary DM can fuse the messages of some groups of DMs into single binary messages using some threshold rule (for example, see [13]).

Given a primary DM and an associated binary decision tree, the consultant DMs will be assigned so that they can assist the primary DM in making the team decision. The suggested way to do this is to assign one consultant DM to each decision node. Then at every decision node t, the primary DM and the corresponding consultant DM will behave like a *two DM tandem team* (a primary DM with a single consultant DM [5], [13], [19]) in order to make a decision for the binary hypothesis testing problem ($\boldsymbol{H}_{t_0}$ vs. $\boldsymbol{H}_{t_1}$). It is important to note that, unfortunately, when composite hypotheses are considered, the conditional independence assumption is lost because, given the true state of the environment, each and every DM can update his/her "beliefs" on the probability distributions of the other DMs. (To see this, consider the composite hypothesis which consists of gaussian elementary hypotheses with different means and very small variance; then, knowledge of the true hypothesis clearly gives specific information on the probable values of this observation.) As we are trying to design an approximate decision scheme, each two DM tandem team is analyzed as though the conditional independence assumption was still valid. The two DM tandem team that *performs* binary hypothesis testing has been extensively analyzed [13]; the three decision thresholds (two for the primary DM and one for the consultant DM) which completely define the optimal team decision making process are as follows :

For the primary DM :

$$\text{If } u_c = 0: \qquad \Lambda(y_p) \underset{u_p=0}{\overset{u_p=1}{\gtrless}} \frac{1-P_F^c}{1-P_D^c}\eta \equiv \eta_0 \tag{11}$$

$$\text{If } u_c = 1: \qquad \Lambda(y_p) \underset{u_p=0}{\overset{u_p=1}{\gtrless}} \frac{P_F^c}{P_D^c}\eta \equiv \eta_1 \tag{12}$$

For the consultant DM :

$$\Lambda(y_c) \underset{u_c=0}{\overset{u_c=1}{\gtrless}} \frac{P_F^1 - P_F^0}{P_D^1 - P_D^0}\eta \equiv \eta_c \tag{13}$$

where :

$$\Lambda(y_n) = \frac{P(y \mid F^1)}{P(y \mid H^0)} \tag{14}$$

is the likelihood ratio of the observation of DM n, and η is defined in Eq. (4); with P_D^i and P_F^i respectively the probability of detection and probability of false alarm for the primary DM when $u_c = i$ was send by the consultant DM ($i = 0, 1$) [5] and, P_D^c and P_F^c respectively the probability of detection and probability of false alarm for the consultant DM.

Then the optimal decision rules for the DMs of the two DM tandem team that tests $\boldsymbol{H}_{t_0}$ versus $\boldsymbol{H}_{t_1}$ can be written by inspection :

For the primary DM :

$$\text{If } u_{c,t} = 0: \qquad \Lambda(y_p) \underset{u_{p,t}=0}{\overset{u_{p,t}=1}{\gtrless}} \frac{1-P_F^{c,t}}{1-P_D^{c,t}}\eta \equiv \eta_0 \tag{15}$$

$$\text{If } u_{c,t} = 1: \qquad \Lambda(y_p) \underset{u_{p,t}=0}{\overset{u_{p,t}=1}{\gtrless}} \frac{P_F^{c,t}}{P_D^{c,t}}\eta_1 \equiv \eta_{1,t} \tag{16}$$

For the consultant DM :

$$\Lambda(y_c) \underset{u_c=0}{\overset{u_c=1}{\gtrless}} \frac{P_F^1 - P_F^0}{P_D^1 - P_D^0}\eta \equiv \eta_c \tag{17}$$

where the notation has the usual meaning and the subscript t indicates that the thresholds are associated with decision node t (i.e., $\boldsymbol{H}_{t_0}$ vs. $\boldsymbol{H}_{t_1}$).

3.4. DISCUSSION

The approximate decision scheme performs especially well as compared to the optimal decision scheme, when there exist particular hypotheses (different for each DM) which a DM can detect better than others. Such DMs are assigned at the decision node of their "expertise", thus reducing the loss of information caused by the processing of their observations into messages, and consequently reducing the degradation of the team performance caused by the approximate decision rules.

Reviewing the decision scheme of the section 3.3, one DM is designated as the primary DM (DM 1 is the primary DM for the team of Figure 1), just like in the optimal decision scheme. Then the multiple hypothesis testing problem is broken down into a sequence of $M-1$ binary hypothesis testing problems with composite hypotheses; this sequence is represented by a decision tree (Figure 2). The consultant DM that is associated with every such binary problem and the primary DM operate as a two DM tandem team. Every time the primary DM makes a decision, the process effectively moves into a new node in the decision tree and a new consultant DM comes to assist the primary DM. The new two DM tandem team will make its decision based only on the observations of its two DMs. Thus, it is assumed that the primary DM does not take explicitly into consideration the messages that he/she may have previously received from other consultant DMs; these messages are taken implicitly into consideration by the arrival of the decision making process at that particular decision node. When the process reaches a terminal node of the tree, the hypothesis corresponding to that terminal node is declared to be true as the final team decision.

5. The two messages assigned to the consulting DM do not have to be denoted 0 and 1. For that matter they can be denoted m_1 and m_2. Without loss of generality assume that :

$$\frac{P(u_c = m_1 \mid H_0)}{P(u_c = m_1 \mid H_1)} \geq \frac{P(u_c = m_2 \mid H_0)}{P(u_c = m_2 \mid H_1)}$$

Then it can be shown that when the primary DM receives $u_c = m_1$, it will always be more likely to decide $u_p = 0$ and when he receives $u_c = m_2$, it will always be more likely to decide $u_p = 1$. Hence the interpretation of m_1 as 0 and of m_2 as 1.

According to the approximate decision scheme, all of the consultant DMs are not taken into consideration in every decision. Therefore, this scheme ignores some of the available information, and hence sacrifices some performance, but in the same time offers two main advantages. First, as was already mentioned, the computational complexity is considerably reduced; in fact by choosing the decision tree appropriately, the on-line complexity can be reduced to $O(\log_2 M)$ and thus the decision process is sped up. Second, each of the consultant DMs improves his/her performance by adjusting his/her binary message to the particular binary detection problem; that is, each consultant DM concentrates on a particular binary problem and gives a better "expert" opinion for (or more applicable to) that problem.

The decision scheme can be characterized as *quasi-sequential.* It is not sequential in the traditional sense which wants DMs to have the options to stop and decide in favor of some hypothesis, or to receive more information. But, it is sequential in the sense that the final team decision is reached with a sequence of preliminary decisions. At every step the primary DM receives a *controlled* new observation (decision) from the appropriately selected consultant DM, but does not receive a new (personal) observation; moreover, at every sequential step the primary DM is faced with a *different* binary hypothesis testing problem.

The approximate decision scheme has been presented and the optimal decision rules for the members of the team have been derived. Then three important questions, which deal with the optimal team architecture and configuration, naturally arise from the above discussion :

(1) *Which of the DMs of the team should be designated as the primary DM?*

(2) *How can the optimal binary decision tree be constructed?*

(3) *Which consultant DM should be assigned to each decision node of the tree?*

The following three subsections attempt to address the above questions.

3.5. SELECTING THE PRIMARY DM

In [13], problems of small cooperative organizations that perform binary hypothesis testing were examined. The objective there was again to keep the combinatorial complexity under control; thus, the difficulties of the problems arose only from the intrinsic complexity of the distributed problems. Different architectures for these organizations were analyzed in a quantitative manner and their performance was compared. Some "common sense" and "intuitively appealing" beliefs were investigated to determine whether they are indeed always true in this framework.

One of the problems examined was to determine the optimal configuration for the tandem team (i.e., determine which DM should be made the primary one). If one DM is better than the other, it is intuitively appealing that the better DM be made the primary DM. Given two DMs one would expect to have the better DM make the team decision, independent of the prior probabilities and the cost assignments. If this was the case, then the optimal way of organizing two DMs would not change, say, as the prior probabilities of the underlying hypotheses vary. This has also been supported with explanations on data compression [5] and with numerical results [19]. But, it is not true in general; it was shown that the optimal configuration depends on parameters external to the team, namely the prior probabilities, on the cost assignments and, in a counterintuitive manner, on the number of messages which the consulting DM is allowed to transmit to the primary DM.

Therefore, it should be evident that no 'globally optimal' primary DM exists for all possible prior probabilities and costs. Consequently, an intuitive and logical approximate solution is to employed that is not demanding computationally. Furthermore, it was also shown in [13] that assigning the best DM as the primary DM leads to very good performance (optimal in most cases and very close to optimal in the rest). Thus, consider the DM with the minimum individual normalized probability of error (or an approximation to it, if the calculation of actual error probability is too demanding), for the given priors and costs, and designate that DM to be the primary DM. This choice for the primary DM is 'robust' in the sense that it leads to good performance, as compared to the optimal, even for a bad choice of the decision tree and consultant DMs.

3.6. CONSTRUCTING THE BINARY DECISION TREE

There exists an exponential number of binary decision trees and there are several approximate ways of selecting one for the problem. We will use a corrected variant of the algorithm presented in [4].

Define the following set which consists of the $[M(M-1)]/2$ possible permutations of pairs of hypotheses (pairs with identical elements are not included) :

$$\boldsymbol{H}^2 = \{(H_m, H_q) | m = 0, 1, ..., M-2, \text{ and } q = m+1, m+2, ..., M-1\} \quad (18)$$

Consider the DM who has been selected to be the primary DM; the observation y_1 of the primary DM is distributed with density $P(y_1|H_m)$ under $H_m (m = 0, 1, ..., M-1)$. Then assign to each pair (H_m, H_q) the distance $d(H_m, H_q, \pi_m, \pi_q)$, where $d(\cdot)$ is some selected stochastic distance measure (for example, the variational distance or the J-divergence; see [9]) that is weighted by the π_m's, the normalized prior probabilities of H_m defined in Eq. (6). Note that a small stochastic distance measure indicates similarity (i.e., closeness) between two hypotheses; hence, a small distance measure implies a difficult decision (i.e., a decision with large error probability). Suppose that $M > 2$ [6]; the decision tree is created as follows [7] :

ALGORITHM.

STEP k ($k = 1, 2, ..., [M(M-3)+4]/2$). Consider the pair of hypotheses with the k-th smallest distance. Then, there exist three possible cases :

CASE I. If neither hypothesis of the pair belongs to the feasible set of a previously defined decision node, then define two terminal nodes at the lowest level of the tree and assign one of the two hypotheses to each. Connect the two terminal nodes by arcs to a new decision node at the second level of the tree; direct the arcs towards the terminal nodes. By definition, the feasible set of the newly generated decision node consists of the union of the feasible sets of the two terminal nodes; that is, it consists of the two hypotheses of the pair with the k-th smallest distance.

CASE II. If only one hypothesis H^* of the pair belongs to the feasible set of a previously defined decision node, denote by t^* the node whose feasible set $\boldsymbol{F}(t^*)$ contains H^* and has the maximum cardinality. Then, define a new terminal node at the same level as t^* and assign H^* to it. Connect the newly defined terminal node and t^* by arcs to a new decision node at the next higher level of the tree; direct the arcs towards the new terminal node and towards t^*. By definition, the feasible set of the new decision node is $\boldsymbol{F}(t^*) \cup \{H^*\}$; if this set has cardinality M then stop.

CASE III. If both hypotheses of the pair belong to the feasible set of some previously constructed decision node(s), then there exist two possibilities. If both hypotheses of the pair belong to the *same* feasible set of some previously defined decision node, then go to the next step $k+1$. Otherwise, if both hypotheses of the pair do *not* belong together to any feasible set of a previously defined node, denote by $t_0(t_1)$ the node of the feasible set with the maximum cardinality that contains the first (second) hypothesis of the pair; connect these two nodes with an arc to a new decision node t at a level that is one higher than the maximum level of t_0 and t_1; direct the arcs towards t_0 and towards t_1. Again, the feasible set of t consists of the union of the feasible sets of t_0 and t_1. If the feasible set has cardinality M, then stop.

The last node generated by the above algorithm is the source node. If a new decision node is defined at the k-th step of the algorithm, then the feasible set of the newly generated decision node contains the two hypotheses having the k-th smallest distance. The decision tree is generated in such a way so as to postpone the most difficult decisions (i.e., decisions between pairs of hypotheses with small distances). We correctly assume that a 'close' error will be less costly, since in the determination of the distance between the hypotheses not only the probability density functions, but also the prior probabilities and the costs were considered (through the normalized prior probabilities).

6. If $M = 2$ the construction of the tree is trivial.

7. Note that the algorithm described in [4] is not correct; it needs some correction so as to avoid a deadlock.

EXAMPLE. CONSTRUCTING THE BINARY DECISION TREE

A seven hypothesis example is presented to demonstrate the construction of a binary decision tree. Before considering the symmetric distance matrix of Table 1, note that it does not have to necessarily be symmetric because the distance measure may not be symmetric (for example, I-divergence).

The algorithm proceeds as follows to generate the decision tree of Figure 2 :

STEP 1. $d(H_0, H_1) = 1$; D_1 is created; $\boldsymbol{F}(D_1) = \{H_0, H_1\}$.

STEP 2. $d(H_2, H_3) = 2$; D_2 is created; $\boldsymbol{F}(D_2) = \{H_2, H_3\}$.

STEP 3. $d(H_5, H_6) = 3$; D_3 is created; $\boldsymbol{F}(D_3) = \{H_5, H_6\}$.

STEP 4. $d(H_1, H_3) = 4$; D_4 is created;
$\boldsymbol{F}(D_4) = \{H_0, H_1, H_2, H_3\}$.

STEP 5. $d(H_0, H_3) = 5$; go to *Step 6*.

STEP 6. $d(H_4, H_5) = 6$; D_5 is created; $\boldsymbol{F}(D_5) = \{H_4, H_5, H_6\}$.

STEP 7. $d(H_4, H_6) = 7$; go to *Step 8*.

STEP 8. $d(H_0, H_2) = 8$; go to *Step 9*.

STEP 9. $d(H_2, H_4) = 9$; S is created;
$\boldsymbol{F}(S) = \{H_0, H_1, H_2, H_3, H_4, H_5, H_6\}$: STOP.

In Figure 2, the terminal nodes are white, the decision nodes are gray and the source node is black. As expected six (6) terminal nodes generate five ($5 = 6 - 1$) decision nodes. The interested reader should note that the algorithm of [4] would brake down because at *Step 4* node D_4 would not be constructed and the algorithm would not recover.

Table 1. – Distance Matrix for the Binary Decision Tree Example

	H_0	H_2	H_2	H_3	H_4	H_5	H_6
H_0		1	8	5	10	10	10
H_1	1	–	10	4	10	10	10
H_2	8	10	–	2	9	10	10
H_3	5	4	2	–	10	10	10
H_4	10	10	9	10	–	6	7
H_5	10	10	10	10	6	–	3
H_6	10	10	10	10	7	3	

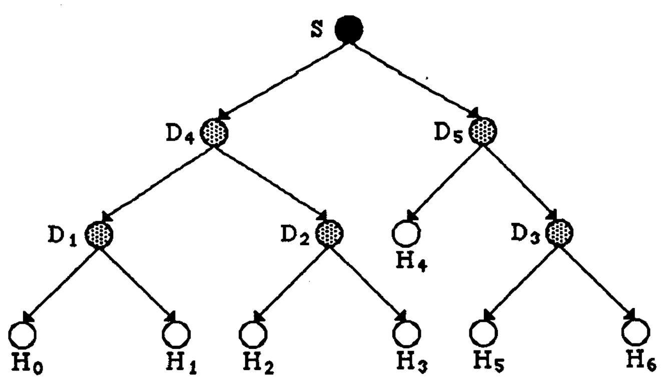

Figure 2. – The Binary Decision Tree of the Example.

3.7. ASSIGNING THE CONSULTANT DMs

In the previous two sections, an algorithm for selecting a primary DM and, given a primary DM, an algorithm for reducing a multiple hypothesis testing problem into a sequence of binary hypothesis testing problems have been presented. Given a primary DM and a binary decision tree, it remains to be seen how should the consultant DMs be assigned to the decision nodes of the tree.

The assignment problem is formulated as a *maximum weight bipartite matching problem* [8]. The two sets of nodes of the bipartite matching problem consist of the $M-1$ consultant DMs and of the $M-1$ non-terminal nodes of the tree. Each consultant DM n is connected to every decision node t with an arc which has an associated weight $w_{n,t} \leq 0$; thus the maximum weight matching will also be a *maximum* matching (i.e., all the nodes will be matched), since each consultant DM will be matched with a non-terminal node and vice-versa. How should the weights $w_{n,t}$ be selected so that the solution to this matching problem yield a good solution of the original M-ary hypothesis testing problem?

Consider some non-terminal node t of the decision tree. There are three factors that need to be considered for a successful assignment of a consultant DM to the binary problems of node t. The first factor is the conditional normalized probability of error that the consultant DM incurs when he/she solves the problem represented by t, conditional on the true hypothesis belonging to the feasible set of t. It is conditional because by the definition of the approximate decision scheme, the consultant DM at decision node t makes his/her decision conditional on the true hypothesis being on the feasible set of t (Eq. (17)). Furthermore, if an error has already occurred in the decision process, the message of the consultant DM, who is supposed to address the wrong problem, is not going to significantly alter the performance of the team (recall that as the decision process moves down the decision tree, the hypotheses become "more similar" both in probability and in cost). Clearly, the smaller the normalized probability of error of a consultant DM the more suitable the DM is for that binary problem.

The second factor that needs to be considered is the individual performance of the primary DM at each decision node t (i.e., the primary DM's normalized probability of error conditional on one of the hypotheses in the feasible set of t being true). It is obvious that the better the performance of the primary DM, the less the need for a consultant DM and vice-versa. Denote by $P^e_{1,t}$ the normalized probability of error (or its approximation) of the primary DM 1 at the decision node t (i.e., $\boldsymbol{H}_{t_0}$ vs. $\boldsymbol{H}_{t_1}$), conditional on the true hypothesis being on the feasible set of t.

The third factor is the prior probabilities and the detection costs. Consider some DM n who can be associated with either of two decision nodes; decision node t or decision node s. Furthermore, assume that DM n has the exact same performance (i.e., normalized probability of error) at both decision nodes t and s, conditional on the true hypothesis being in the feasible set of t and s respectively. Other things being equal, on which of the two decision nodes should DM n be assigned?

DM n should be associated with the decision node whose feasible set has the *larger sum of elementary normalized prior probabilities*. This is easier to comprehend in the case of the minimum error cost function; in this case, DM n should be associated with the decision node whose feasible set has the larger sum of prior probabilities. This should be intuitive as, by being associated with the decision node whose feasible set has the larger sum of prior probabilities, DM n's decision will be taken into account by the team more often.

Thus, the weight $w_{n,t}$ of assigning the consultant DM n to the binary hypothesis problem of a non-terminal node t of the decision tree is defined as follows :

$$w_{n,t} = (1 - P^e_{n,t})P^e_{1,t}\left[-\log\left(1 - \sum_{m\in F(t)} \pi_m\right)\right] \quad (19)$$

Remark 1. Clearly : $w_{n,t} \leq 0$.

Remark 2. The logarithm in Eq.(19) is included to emphasize the assignment of the better DMs to the ('important') non-terminal nodes which are high on the tree (because these nodes have large normalized prior probability π_t). In fact, note that since for the source node $\pi_s = 1$, the consultant DM with the smallest normalized probability of error for the source node's binary problem (i.e., $\boldsymbol{H}_{s_0}$ vs. $\boldsymbol{H}_{s_1}$) will *always* be assigned there.

Remark 3. There exist several algorithms to solve the maximum weight bipartite matching problem which require $O(M^3)$ time [10].

Remark 4. As has already been mentioned, the (restrictive) implicit assumption in the definitions of the normalized prior probabilities for both elementary and composite hypotheses (Eqs.(5) and (8)) is that, given the true hypothesis, the relative frequencies of the different types of error are given by the prior probabilities. In the decision scheme that was presented, the M-ary hypothesis testing problem was broken down into a sequence of composite binary hypothesis testing problems; the composite hypotheses consist of *similar* hypotheses. Thus even though the assumption is not true in general, it is quite reasonable for the proposed decision scheme especially if the probability density functions are smooth.

Remark 5. To compute the decision thresholds a Gauss-Seidel algorithm similar to the one described in [23] should be employed. As stated earlier, the algorithm is guaranteed to converge to decision rules which satisfy the necessary optimality conditions, but there are no guarantees that these will actually be the optimal decision rules. Still, in the binary hypothesis testing case there exists a single likelihood ratio for each DM; therefore, we can perform one dimensional search and guarantee that the algorithm converges to the optimal decision rules. This is an additional computational advantage of the approximate decision scheme.

8. For an in depth treatment of *combinatorial optimization* the reader is refered to [10].

3.8. IMPROVING ON THE APPROXIMATE DECISION SCHEME

The approximate decision scheme described above employs common sense and intuition, together with some straightforward mathematics, to yield a very good approximate solution for the very complicated *Problem* of section 2. Because of the several subjective and arbitrary choices that are made (selection of the primary DM, construction of the binary decision tree and the matching problem), it is virtually impossible to obtain analytically any meaningful bounds for the performance of the decision scheme; this could be an interesting topic for future research.

We do not claim to have derived the "best" approximate solution. On the contrary there exist several ways to improve it, but our objective was just to demonstrate that relatively simple procedures may be designed for successfully tackling complicated problems. Indeed, the performance of the decision scheme can be improved, without forbidding increases in the computational complexity being induced if, for example, one or more of the following factors are taken into consideration :

1. A better algorithm for the generation of the binary decision tree should be derived. This should take into account its particular application in our problem. For example, in the proposed algorithm the pair of hypotheses with the kth smallest distance is considered at each step k. Once two elementary hypotheses are to be combined into a composite hypothesis, the probability distribution of the composite hypothesis consists of a weighted convolution of the distributions of the elementary hypotheses. Therefore, it should be preferable at every step of the algorithm to calculate a new distance matrix by taking explicitly into consideration the newly generated composite hypothesis, instead of always using the initial distance matrix (i.e., the distance matrix between the elementary hypotheses).

2. A more sophisticated selection of the weights for the matching problem.

3. Relevant results from information theory should be employed in this framework.

4. For the decision of a particular consultant DM to be considered by the primary DM, a particular sequence of decisions must be made first. For example, for the decision of the consultant DM corresponding to decision node D_2 (H_2 vs. H_3) to be considered, the team must decide 0 at the source node S and decide 1 at decision node D_4. This provides some additional information to the consultant DM corresponding to D_2; he/she can therefore update his/her beliefs of the distribution of the observation of the primary DM, in order to produce an even better decision.

Furthermore, the most important advice for improving the performance of the approximate decision scheme is to try to take maximum advantage of the particular structure of the problem. We should carefully examine the characteristics of the problem (i.e., DMs, hypotheses, costs, etc.) and try to assist the 'mathematical solutions' by making educated choices, which the (simple, approximate) analysis may overlook. Keeping this in mind, we present the following specific example :

EXAMPLE. A MULTIATTRIBUTE HYPOTHESIS TESTING PROBLEM

Consider a particular instance of the *Problem* of section 2, in which each of the multiple hypotheses consists of a set of independent or loosely dependent attributes. For example, suppose that the multiple hypotheses describe the characteristics of a refrigerator; the size, the color, the weight, the temperature and the motor of the refrigerator are almost independent attributes (the size of a refrigerator contains almost no information about its color and temperature), but knowledge of all of them can provide considerable information on determining the its exact model and make.

Moreover, suppose that the observation of each of the consultant DMs contain information on just one of the attributes. In the refrigerator example, one consultant DM has a tape measure and can measure its size, another has a chart and can determine its exact color, another has a scale and can weight it, another has a thermometer and another has a voltmeter and an ampmeter. On the other hand, the observation of the primary DM contains some information about all the attributes. The objective is to design a decision protocol that will optimize the team performance.

The problem just described fits the framework of the more general problem introduced in section 2 since the team performs multiple hypothesis testing and consists of a single primary DM, and several consultant DMs. We are going to argue that the additional structure imposed on the original problem leads to very good and considerably simpler solutions.

The observation of only one of the DMs contains information on all the attributes of the hypotheses; this DM presents us with a clear-cut choice for the primary DM of the team. Furthermore, since each of the consultant DMs is an 'expert' in just one of the attributes of the hypotheses, he/she is immediately 'matched' with a particular subproblem, and can use his/her messages to transmit to the primary DM information about just this particular attribute without any degradation in the team performance. Therefore, the optimal configuration of the team is determined by inspection. Finally since the various attributes of the hypotheses are independent or loosely dependent, the primary DM and the appropriate consultant DM can indeed perform as a two DM tandem team because the conditional independence assumption will not be violated.

Hence, the only remaining issue that needs to be resolved involves the sequence with which the decisions are made; that is, it involves the construction of the decision tree (Figure 3). The decision tree will be a little different than the one described in the previous section. First, it does not have to be binary; for example, if the refrigerator can have one of three different colors, it should be worthwhile to endow the 'color expert' with three messages, one for each color.

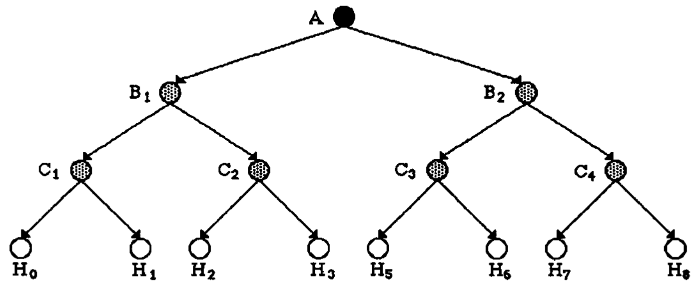

Figure 3. – Decision Tree for Three-Attribute Problem.

Second, since the attributes are independent with each other, the depth of the decision tree will be constant (since knowledge of one attribute does not contain any information about another, all the attribute combinations are possible).

Third, note that in this type of problem every consultant DM is associated with an attribute of the multiple hypothesis, and is not associated with a composite hypothesis testing problem. Consequently, each consultant DM will be associated with *all the decision nodes at a given level of the decision tree* and will not be associated with just one decision node (for the decision tree of Figure 3, the consultant DM for the first attribute is associated with node A, the consultant DM for the second (third) attribute is associated with all the B (C) nodes). To see this, consider again the refrigerator example and suppose that its attributes are examined in the given order : size, color, weight, etc.. In the third stage of the game when the weight of the refrigerator is examined (the C nodes in Figure 3), *independent of the team decisions up to that point* (i.e., the decisions on size and color), the two DM tandem team consisting of the primary DM and the 'weight expert' consultant DM will always have to decide on the weight of the refrigerator. Thus, the 'weight expert' is associated with all the decisions that take place in the third level (from above) of the decision tree.

Therefore, to construct the decision tree only a testing sequence for the attributes of the multiple hypotheses has to be generated. To do this, we employ some stochastic distance measure and arrange the sequence of the decisions in such a way, so that the more difficult decisions are postponed for later (i.e., the lower levels of the tree).

4. Numerical Examples

Two numerical examples follow to demonstrate the benefits of the approximate decision scheme. The organization consists of three DMs, denoted A, B and C. There are three different possible states for the environment (i.e., hypotheses); hypothesis H_m occurs with an a priori probability $P(H_m$, for $m = 0, 1, 2$. The objective of the organization is to detect the value of a signal; under hypothesis H_m the value of the signal is μ_m (the same one for all three DMs). The observation of DM n ($n = A, B, C$) is corrupted by a zero mean gaussian noise with variance σ_n^2. The minimum probability of error cost function is employed; that is :

$$J(u, H_m) = \begin{cases} 1 & \text{if} : u \neq m \\ 0 & \text{if} : u = m \end{cases} \quad \text{for} : u, m = 0, 1, 2. \tag{20}$$

Each consultant DM communicates to the primary DM one of three possible messages (call them 0, 1 and 2). Three computer programs were developed; one obtains the optimal decision rules of Eqs.(1)-(3), one determines the decision rules of the approximate decision scheme, and one solves the centralized version of the problem.

EXAMPLE 1.

The possible values for the signal are : $\mu_0 = -10$, $\mu_1 = 0$ and $\mu_2 = 10$. The prior probabilities of the hypotheses are : $P(H_0) = 0.2$, $P(H_1) = 0.3$ and $P(H_2) = 0.5$. The variances of the observations of the DMs are : $\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = 100$. We decided to assign equal variances to all three DMs in the first example so as to avoid the problem of determining the best configuration for the organization.

The optimal decision rules for the two consultant DMs are :

$$u_B = \gamma_B(y_B) = \begin{cases} 2; & y_B > 6.1670 \\ 1; & 6.1670 > y_B > -8.7463 \\ 0; & -8.7463 > y_B \end{cases} \tag{21}$$

$$u_C = \gamma_C(y_C) = \begin{cases} 2; & y_C > 4.5424 \\ 1; & 4.5424 > y_C > -9.3605 \\ 0; & -9.3605 > y_C \end{cases} \tag{22}$$

Even though both consultant DMs are *identical*, their optimal decision rules are *not* identical (or even symmetric). This is further confirmation of the results in [24] (for exponentially distributed observations) and in [26] (for discrete observations). The optimal decision rules of the primary DM are given by :

$$u_A = \gamma_A(y_A, u_B = 1, u_C = 1)$$
$$= \begin{cases} 2; & y_A > 37.1381 \\ 1; & 37.1381 > y_A > 12.5427 \\ 0; & 12.5427 > y_A \end{cases} \tag{23a}$$

$$u_A = \gamma_A(y_A, u_B = 1, u_C = 2)$$
$$= \begin{cases} 2; & y_A > 24.5722 \\ 1; & 24.5722 > y_A > -0.6766 \\ 0; & -0.6777 > y_A \end{cases} \tag{23b}$$

$$u_A = \gamma_A(y_A, u_B = 1, u_X = 3)$$
$$= \begin{cases} 2; & y_A > 10.4742 \\ 1; & 10.4742 > y_A > -13.4110 \\ 0; & -13.4110 > y_A \end{cases} \tag{23c}$$

$$u_A = \gamma_A(y_A, u_B = 2, u_C = 1)$$

$$= \begin{cases} 2; & y_A > 24.0025 \\ 1; & 24.0025 > y_A > -1.1176 \\ 0; & -1.1176 > y_A \end{cases} \quad (23d)$$

$$u_A = \gamma_A(y_A, u_B = 2, u_C = 2)$$
$$= \begin{cases} 2; & y_A > 11.4365 \\ 1; & 11.4365 > y_A > -14.3380 \\ 0; & -14.3380 > y_A \end{cases} \quad (23e)$$

$$u_A = \gamma_A(y_A, u_B = 2, u_C = 3)$$
$$= \begin{cases} 2; & y_A > -2.6614 \\ 1; & -2.6614 > y_A > -27.0713 \\ 0; & -27.0713 > y_A \end{cases} \quad (23f)$$

$$u_A = \gamma_A(y_A, u_B = 3, u_C = 1)$$
$$= \begin{cases} 2; & y_A > 9.9512 \\ 1; & 9.9512 > y_A > -14.2765 \\ 0; & -14.2765 > y_A \end{cases} \quad (23g)$$

$$u_A = \gamma_A(y_A, u_B = 3, u_C = 2)$$
$$= \begin{cases} 2; & y_A > -2.6147 \\ 1; & -2.6147 > y_A > -27.4970 \\ 0; & -27.4970 > y_A \end{cases} \quad (23h)$$

$$u_A = \gamma_A(y_A, u_B = 3, u_C = 3)$$
$$= \begin{cases} 2; & y_A > -16.7126 \\ 1; & -16.7126 > y_A > -40.2302 \\ 0; & -40.2302 > y_A \end{cases} \quad (23i)$$

These decision rules result in an expected cost (i.e., probability of error) of 0.2558.

The proposed approximate reduces the problem into two binary hypothesis testing subproblems. The first subproblem tests compares hypotheses $\boldsymbol{H}_{\boldsymbol{F}_0}$ and $\boldsymbol{H}_{\boldsymbol{F}_1}$, where $\boldsymbol{F}_0 = \{0, 1\}$ and $\boldsymbol{F}_1 = \{2\}$; that is, it compares H_2 with the composite hypothesis of H_0 and H_1. (Note that H_1 is stochastically closer to H_0 than to H_2 because, while the absolute magnitude of the difference in the means of H_1 and H_0 and of H_2 and H_1 is equal, their prior probabilities are closer.) The decision rules are for the consultant DM :

$$u_B = \gamma_B(y_B) = \begin{cases} 1; & y_B > 0.7658 \\ 0; & 0.7658 > y_B \end{cases} \quad (24)$$

and for the primary DM :

$$u_{A,1} = \gamma_{A,1}(y_A, u_B = 0)$$
$$= \begin{cases} 1; & y_A > -4.1999 \\ 0; & -4.1999 > y_A \end{cases} \quad (25a)$$

$$u_{A,1} = \gamma_{A,1}(y_A, u_B = 1)$$
$$= \begin{cases} 1; & y_A > 13.9842 \\ 0; & 13.9842 > y_A \end{cases} \quad (25b)$$

The second subproblem compares H_1 to H_0. The decision rules are for the consultant DM :

$$u_C = \gamma_C(y_C) = \begin{cases} 1; & y_C > -6.6243 \\ 0; & -6.6243 > y_C \end{cases} \quad (26)$$

and for the primary DM :

$$u_{A,2} = \gamma_{A,2}(y_A, u_C = 0)$$
$$= \begin{cases} 1; & y_A > -16.1273 \\ 0; & -16.1273 > y_A \end{cases} \quad (27a)$$

$$u_{A,2} = \gamma_{A,2}(y_A, u_C = 1)$$
$$= \begin{cases} 1; & y_A > 6.9401 \\ 0; & 6.9401 > y_A \end{cases} \quad (27b)$$

These result in a probability of error of 0.2918. Therefore, the deterioration in the team performance if the approximate decision scheme is employed is 14.1%. On the other hand, the approximate decision scheme reduces the computational requirements by more than 60%. In fact, the optimal distributed decision rules for the organization with ternary communication messages are defined by twenty two thresholds (two for each of the consultant DMs, and two for the primary DM for each of the nine possible sets of messages that can be received from the consultant DMs), while the approximate decision scheme requires the determination of only six thresholds (one for each consultant DM and two for the primary DM for each of the two subproblems). It should be clear that the approximate decision scheme is computationally much more efficient.

The optimal decision rules for the centralized case are :

$$\overline{u} = \overline{\gamma}(y_A, y_B, y_C) = \begin{cases} 2; & \overline{y} > 3.2972 \\ 1; & 3.2972 > \overline{y} > -6.3516 \\ 0; & -6.3516 > \overline{y} \end{cases} \quad (28)$$

where :

$$\overline{y} = \frac{\sigma_B^2\sigma_C^2 y_A + \sigma_C^2\sigma_A^2 y_B + \sigma_A^2\sigma_B^2 y_C}{\sigma_B^2\sigma_C^2 + \sigma_C^2\sigma_A^2 + \sigma_A^2\sigma_B^2} \quad (29)$$

These result in a probability of error of 0.2404. Therefore, the optimal solution for the distributed problem is within 6.6% of the centralized solution. This is another confirmation for the conclusion of [17] that, in the distributed hypothesis testing environment, a small number of communication messages (a number comparable to the number of hypotheses) is sufficient for the distributed organization to perform very well when compared to the centralized counterpart.

EXAMPLE 2.

The possible values for the signal are : $\mu_0 = -5$, $\mu_1 = 0$ and $\mu_2 = 10$. The prior probabilities of the hypotheses are : $P(H_0) = 0.3$, $P(H_1) = 0.4$ and $P(H_2) = 0.3$. The variances of the observations of the DMs are : $\sigma_A^2 = 50, \sigma_B^2 = 100, \sigma_C^2 = 150$. The explicit decision rules are not presented because they do not provide any additional insight to the problem.

The optimal solution requires that DM A be the primary DM and results in a probability of error of 0.34103

The approximate decision scheme again reduces the problem into two binary hypothesis testing subproblems. In the first subproblem the team decides between H_2 and the composite hypothesis of H_0 and H_1. In the second subproblem, the team has to decide between H_0 and H_1.

The three DMs receive observations corrupted by zero mean gaussian noise. The variance of DM A is the smallest and the variance of DM C is the largest. Therefore, DM A and DM C are the best and worst DMs respectively. Consequently, DM A is designated as the primary DM, DM B is designated as the consultant DM for the first subproblem and DM C is designated as the consultant DM for the second subproblem. The resulting probability of error is 0.36469.

The deterioration in the performance of the team if the approximate decision rules are employed is 6.9% of the optimal, but the reduction in the computational requirements is more than 50% (assuming that the correct primary DM is known a priori; otherwise the reduction in the computational requirements jumps to 85%).

The probability of error for the centralized case is 0.3325. Therefore, the optimal solution is within 2.6% of the centralized solution, again confirming that ternary communication messages result in very good performance in ternary hypothesis testing.

5. Summary

As the number of the DMs, or of the hypotheses, or of the messages increases, the combinatorial complexity of problems in this framework increases exponentially. In order to keep the complexity under control, approximate solutions need to be developed. Since we also desire that the approximate decision schemes achieve good performance, the particular characteristics of the problem should be taken into account. This leads to *problem dependent analysis* which, we believe, should be the focus of future research. To demonstrate this approach, we discussed the issues involved with the reduction of a multiple hypothesis testing problem into a sequence of simpler hypothesis testing problems, under different operating conditions. Numerical examples were used to show that the deterioration in performance if the proposed approximate decision rules are implemented is more than offset by the reduction in the combinatorial complexity of the problem.

REFERENCES

[1] Chair, Z., and P.K. Varshney, "Optimum Data Fusion in Multiple Sensor Detection Systems", *IEEE Trans. on Aerospace and Electronic Systems*, AES-22 :1, 1986, pp. 98-101.

[2] Chair, Z., and P.K. Varshney, "Distributed Bayesian Hypothesis Testing with Distributed Data Fusion", *IEEE Trans. on Systems, Man and Cybernetics*, SMC-18 :5, 1988, pp. 695-699.

[3] Dasarathy, B.V., *Decision Fusion*, IEEE Computer Society Press, Los Alamitos, CA, 1994.

[4] Dermidas, K., "Distributed Sensor Data Fusion with Binary Decision Trees," *IEEE Trans. on Aerospace and Electronic Systems*, AES-25 :1, 1989, pp. 643-649.

[5] Ekchian, L.K., and R.R. Tenney, "Detection Networks", *Proceedings of the 21st IEEE Conference on Decision and Control*, 1982, pp. 686-691.

[6] Han, J., Varshney, P.K., and R. Srinivasan, "Distributed Binary Integration," *IEEE Trans. on Aerospace and Electronic Systems*, AES-29 :1, 1993, pp. 2-8.

[7] Hoballah, I.V., and P.K. Varshney, "Neyman-Pearson Detection with Distributed Sensors", *Proceedings of the 25th IEEE Conference on Decision and Control*, 1986, pp. 237-241.

[8] Hoballah, I.Y., and P.K. Varshney, "Distributed Signal Detection," *IEEE Trans. on Information Theory*, IT-35 :5, 1989, pp. 995-1000.

[9] Kazakos, D., and P. Papantoni-Kazakos, *Detection and Estimation*, Computer Science Press, New York, NY, 1990.

[10] Papadimitriou, C.H., and K. Steiglitz, *Combinatorial Optimization : Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[11] Papastavrou J.D., Athans M., "A Distributed Hypothesis-Testing Team Decision Problem with Communications Cost," *System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches*, D. Reidel Publishing Company, 1987, pp. 99-130.

[12] Papastavrou J.D., Athans M., "Distributed Detection by a Large Team of Sensors in Tandem," *IEEE Trans. on Aerospace and Electronic Systems*, AES-28 :3, 1992, pp. 639-653.

[13] Papastavrou J.D., Athans M., "On Optimal Distributed Decision Architectures in a Hypothesis Testing Environment," *IEEE Trans. on Automatic Control*, AC-37 :8, 1992, pp. 1154-1169.

[14] Papastavrou J.D., Athans M., "The Team ROC Curve in a Hypothesis Testing Environment," to appear in the *IEEE Trans. on Aerospace and Electronic Systems*, 1994.

[15] Papastavrou J.D., Nof S.Y., "Decision Integration Fundamentals in Distributed Manufacturing Topologies," *IIE Transactions*, 24 :3, 1992, pp. 27-42.

[16] Polychronopoulos, G., and J.N. Tsitsiklis, "Explicit Solutions for Some Simple Decentralized Detection Problems", *IEEE Trans. on Aerospace and Electronic Systems*, AES-26 :2, 1990, pp. 282-292.

[17] Pothiawala J., Papastavrou J.D., Athans M., "Designing an Organization in a Hypothesis-Testing Environment," *1989 Symposium on Command and Control Research*, Washington D.C., June 1989, pp. 116-123.

[18] Reibman, A.R., and L.W. Nolte, "Optimal Detection and Performance of Distributed Sensor Systems", *IEEE Trans. on Aerospace and Electronic Systems*, AES-23 :1, 1987, pp. 24-30.

[19] Reibman, A.R., and L.W. Nolte, "Design and Performance Comparison of Distributed Detection Networks", *IEEE Trans. on Aerospace and Electronic Systems*, AES-23 :6, 1987, pp. 789-797.

[20] Sadjadi, F., "Hypothesis Testing in a Distributed Environment", *IEEE Trans. on Aerospace and Electronic Systems*, AES-22 :2, 1986, pp. 134-137.

[21] Simon, H.A., *Administrative Behavior : A Study of Decision-Making Processes in Administrative Organization*, 3rd Edition, Free Press, New York, 1976.

[22] Srinivasan, R., "A Theory of Distributed Detection", *Signal Processing*, 11 :4, 1986, pp. 319-327.

[23] Tang, Z.B., et al., "A Distributed M-ary Hypothesis-Testing Problem with Correlated Observations", *Proc. of the 29th IEEE Conf. Dec. and Control*, Piscataway, NJ, 1989, pp. 562-568.

[24] Tenney, R.R. and N.R. Sandell, Jr., "Detection with Distributed Sensors", *IEEE Trans. on Aerospace and Electronic Systems*, AES-17 :4, 1981, pp. 501-510.

[25] Tsitsiklis, J.N., and M. Athans, "On the Complexity of Decentralized Decision Making and Detection Problems", *IEEE Trans. on Automatic Control*, AC-30 :5, 1985, pp. 440-446.

[26] Tsitsiklis, J.N., "Decentralized Detection by a Large Number of Sensors", *Mathematics of Controls, Signals and Systems*, 1, 1988, pp. 167-182.

[27] Tsitsiklis, J.N., "Extremal Properties of Likelihood-Ratio Quantizers", *Proc. of the 29th IEEE Conf. Dec. and Control*, Piscataway, NJ, 1989, pp. 2680-2685.

[28] Tsitsiklis, J.N., "Decentralized Detection", *Advances in Statistical Signal Processing*, vol. 2, 1993, pp. 297-344.

[29] Thomopoulos, S.C.A., et al., "Optimal Decision Fusion in Multiple Sensor Systems", *IEEE Trans. on Aerospace and Electronic Systems*, AES-23 :5, 1987, pp. 644-653.

[30] Thomopoulos, S.C.A., et al., "Optimal Distributed Decision Fusion", *IEEE Trans. on Aerospace and Electronic Systems*, AES-25 :5, 1989, pp. 761-765.

[31] Thomopoulos, S.C.A., "Sensor Integration and Data Fusion," *J. Robotics Systems*, 7 :3, 1990, pp. 337-372.

[32] Thomopoulos, S.C.A., and N.N. Okello, "Distributed Detection with Consulting Sensors and Communications Cost," *IEEE Trans. on Automatic Control*, 37 :9, 1992, pp. 1398-1405.

[33] Viswanathan, R., and S.C.A. Thomopoulos, "Optimal Serial Distributed Decision Fusion", *IEEE Trans. on Aerospace and Electronic Systems*, AES-24 :4, 1988, pp. 366-375.

Manuscrit reçu le 18 mai 1994

L'AUTEUR

J.D. PAPASTAVROU

Jason D. Papastavrou was born in Athens, Greece, in 1962. He received the B.S. degree in mathematics, the M.S. and Ph.D. degrees in electrical engineering all from the Massachusetts Institute of Technology, Cambridge, MA, in 1984, 1986, and 1990, respectively. Since 1990 he has been with the School of Industrial Engineering at Purdue University, West Laffayette, IN, where he is currently Assistant Professor. His research interests are in the areas of transportation problems in operations research, decentralized decision making for manufacturing systems, and warning systems.