JUNE 1989 LIDS-P-1885

DESIGNING AN ORGANIZATION IN A HYPOTHESIS TESTING FRAMEWORK

by

Jason D. Papastavrou, Javed Pothiawala and Michael Athans

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

This paper deals with problems of quantitative organizational design. We endorse the development of a Normative Decision Theory for the quantitative study of distributed organizations. We suggest that Hypothesis Testing is an appropriate paradigm for this framework. To demonstrate this, we present numerical results for a team consisting of two decision makers (DMs) which performs binary hypothesis testing and study the effects of different communication protocols; we show that near-optimal performance can be achieved with few communication bits. We summarize some other organizational design problems we have studied and present some tentative conclusions and suggestions for future research.

1. INTRODUCTION

This paper is an informal progress report on research we have been conducting during the past few years. Our main research goal is to develop basic understanding of the decision making process in distributed organizations. To complement empirical studies, we need to develop a normative decision theory for designing superior organizations. The development of such a theory can be used as a tool to assist decision makers into improving the quality of their decisions, and to provide performance benchmarks in empirical studies that capture the bounded rationality of human decision makers and the often chaotic behavior of human organizations.

To appear en Proc. 1989 Symposium on C^2 Research, Washington, D.C.

To achieve our goals we need a specific paradigm which represents simple decision making, and whose centralized version is easy to formulate, solve and compute; for this we employ the problem of hypothesis testing. We want to design a team to perform binary hypothesis testing (e.g. target detection) using several DMs. Each DM has his own private sensor and computational capability to process his own data. The performance of the DM operating in isolation can be quantified in terms of his Receiver Operating Characteristic (ROC) curve; we remark that ROC curves can be derived using empirical data on human decision makers. To improve performance DMs should not operate in isolation; it is desirable to have many DMs operate as a team. For this we have to define the architecture of the team, establish communication protocols, and design the decision protocols which will achieve near-optimal performance.

We employ a binary hypothesis testing model, which can be generalized to more general hypothesis problems. These are indeed generic in the situation assessment C^2 function. We would like to develop a quantitative methodology to deal with them.

In Section 2 we present the motivation for and the guidelines to a Normative Decision Theory and we explain why the Decentralized Hypothesis Testing Framework was chosen. In Section 3 we examine a particular problem in which we examine the effects of different communication protocols on team performance. In Section 4 we summarize other results we have derived recently and finally in Section 5 we present some thoughts on where research should focus on next.

2. THE NORMATIVE DECISION THEORY

2.1 General Remarks

Decision making research has been the meeting point of psychologists, philosophers, sociologists, economists, organizational theorists, statisticians and, more recently of operational researches and engineers. They all work together trying to develop a theory which will improve the quality of decision making both for an individual and for an organization. Gaining a better understanding of how good decisions are made is an essential prerequisite to the goal of increased productivity and improved allocation of resources.

There are two distinct points of view in modeling decision making, which are reflected in the two schools of thought which have been developed: the *organizational* decision theory and the *behavioral* decision theory, [18]. Behavioral decision theory is essentially cognitive and generally uses experimental methods while organizational decision theory is primarily theoretical and naturalistically oriented, examining rather conspicuous individual and social phenomena.

The need for the development of a Normative Decision Theory emanates from both organizational and behavioral decision theories. This normative decision theory should be seen as a tool which will assist DMs in improving the quality of their decisions. We first need to develop these *normative* decision models and then test them in practice to obtain *descriptive* models which fully reflect the so-called "human bounded rationality.". We then need to combine both into *normative/descriptive* models which will be more accurate and realistic in predicting the actions of DMs. Hopefully, similar results can be obtained for teams of DMs, although the problems will be much more complex.

In our view, the normative decision theory (among other things) should capture the following two processes:

1. the development of a DM to an *expert* DM and,
2. the development of a team of expert DMs to an *expert* team of expert DMs [15].

2.2 The Meaning of Expert

The word expert is often associated with two disjoint, but complementary meanings. The first meaning is a *relative* meaning which is very close to the everyday use of the word. When there are several DMs to perform a certain task, the DM (DMs) who can best perform the task is (are) considered to be experts for that particular task. Thus, we need if at all possible, to compare two DMs capable of doing the same task and state that one is "better" than the other. In the problem of binary hypothesis testing the dominance of one ROC curve over another can be used to quantify such a comparison.

The second meaning which we associate with the word expert is *absolute* and could be surprising. We consider a DM who has to perform a certain task to be an expert if he performs the task to the best of his ability; that is we consider a DM be an expert in this sense if he performs at his potential. Every organization has many DMs who are inherently different and who cannot perform the same tasks with the same degree of success. Simon [14,p.36] asserts that "what a person cannot do he will not do, no matter how hard he wants to do it." Thus, an organization should be overly satisfied by having him perform at his level best.

The normative decision theory should first address the issues of developing decision makers to experts in the absolute sense (i.e. performing at their potential); determining optimal training protocols would fall in this category. Second, it should address the issues of organizing, given certain exogenous constraints, the decision makers as the best possible (expert) team; thus developing the decision makers to experts in the relative sense with respect to the tasks which need to be performed within the team. It should capture the fact that the behavior of a DM can be different in isolation vis-a-vis the participation in a cooperating team. The exogenous constraints could and should not only be quantitative constraints (i.e. constraints on the number and the quality of the decision makers, on the available communication links and technology), but also constraints which take the form of issues of resiliency in cases of failures of decision makers and of failures because of enemy interactions.

Both goals are extremely difficult and complex, but it seems to us that the second is even more difficult than the first because of the large number of different ways in which an organization can be set up. Each domain of the normative decision theory can be researched independently, but only development in both will bring together a complete theory which will in turn result in an optimal organization, that is a true expert team of experts.

In our research we deal with the second goal of the proposed theory: the design of an organization. We assume that the decision makers are experts in the absolute sense and try to develop them into an expert team. We study elementary problems which need to be analyzed, solved and understood if a theory is ever to be developed. We try to construct organizational 'building blocks' to be used in the design of bigger organizations. Unfortunately, some of the results that we reported last year, [11], show that it is very difficult to derive general results on how to organize even small teams of DMs with differing expertise. For example, we showed in [11] that in the context of distributed hypothesis testing, that the conjecture of using the "best" DM (the one with the dominant ROC curve) make the final decision in a tandem team is false. There are special problems in which it is better (by a very small amount as measured by the probability of error) to have the "worse" DM carry out the team decision.

2.3 The Hypothesis Testing Framework

There are two major reasons for choosing the decentralized hypothesis testing framework for this research. First, problems in this framework are very simple to describe so that decision scientists without great mathematical sophistication can understand them and draw conclusions from analyzing their solutions. In fact, we should emphasize that problems in this framework look deceptively simple. Moreover these problems have trivial centralized counterparts which implies that all the difficulties and the deterioration in performance occurred because of the decentralization. We began our research by trying to formally prove the most basic and 'obvious' results which existed in the literature as conjectures and we were surprised by the degree of difficulty for their solution, their inherent complexity and the counterexamples we derived.

The second major reason for employing the decentralized hypothesis testing framework for the development of a normative decision theory is that decentralized hypothesis testing is in itself a very interesting subject which has several applications, especially in the area of target detection, classification, and discrimination in the surveillance function.

We know that problems of this type have been shown to be NP-complete [5],[6]. This result identifies the difficulties we are faced with. But, like in the case of the Travelling Salesman Problem (TSP), these problems have so many important practical applications so that merely identifying the difficulties associated with them is not enough. We have to develop new mathematical techniques to solve them or to at least obtain satisfactory heuristics. As Simon [16] said: "...because of the complexity of the environment, one has but two alternatives: to build optimal models by making simplifying assumptions or to build heuristic models that maintain greater environmental realism."

The purpose of this research is not to demonstrate the difficulties which arise because of the combinatorial explosion; the NP-completeness is a testament to these. In most well known NP-complete problems the difficulties arise only because of the combinatorial explosion. For example it is trivial to optimally solve the TSP for up to four or five nodes. But distributed hypothesis testing problems are different because difficulties arise even in the simplest versions. Hence in order to keep the combinatorial explosion under control we test only a small number of hypotheses (two or three), employ a small number of decision makers (up to three) and use limited communications, so that we concentrate on the difficulties which arise because of the inherent complexity of optimizing the decision rules. Only by understanding these difficulties and overcoming them, we will be able to make educated generalizations in order to build good heuristics and eventually achieve a truly optimal solution.

We do not deal with problems of developing decision makers to experts in the absolute sense. We assume throughout that the decision makers perform at their potential which is fixed and cannot be improved. Following Simon's advice we make great simplifications. In this context the decision makers can be seen as 'perfect decision robots.' That is they do not experience conflicts, defensive avoidance, regrets, anchoring, recency, hopes, aspirations, emotions, coercion, confrontations or any of the conditions which make modeling human decision making so difficult. The decision makers behave like computer processors who follow a code and strive to make decision to minimize the expected team cost. The negative of this expected team cost can be seen as the complete team utility function. This cost/utility function describes completely the whole persona of the decision makers since this is the only thing that influences their decision.

3. IMPACT OF COMMUNICATION ON TEAM PERFORMANCE

3.1 Introduction

The coordination of distributed DMs requires a certain amount of communication. In military problems the amount of communication resources can be costly (due to limited bandwidth, jamming, enemy intercepts, covert operations etc). Thus, limited and costly communications must be considered as an essential ingredient of organizational design and performance.

Many problems have been studied in a framework similar to the one we are going to employ ([2]-[12]) and many results were obtained and conjectures (generalizations) drawn. We consider problems in which the decision makers are given and the optimum architecture for the organization satisfying certain requirements is requested. The environment consists of discrete hypotheses which occur with prior probabilities known to all the decision makers. The decision makers receive noisy observations of the environment. Then preliminary decisions are made and communications take place until a final team decision is reached. The team incurs a cost which depends on the final team decision and on the true hypothesis. The costs are known a priori by all the team members.

3.2 Problem Formulation

A team consisting of two DMs has to distinguish between two hypotheses based upon uncertain measurements (Figure 1). The objective of the team is to minimize the probabilitity of error associated with the final team decision.

Each DM receives one noisy observation; these are assumed to be conditionally independent. Each DM has his own capability described by his ROC curve [11]. DM A, the *consulting* DM, makes a decision based on his own measurement and transmits it to DM B, the *primary* DM. Then DM B makes the final team decision based on his own measurement and on the communication from DM A. In the general case DM A can transmit one of K messages. In the simplest case K=2 and only one bit is required; this can be interpreted as the tentative binary decision of DMA. The next simplest case is K=3 which corresponds to 1.5 bits, and so on. The optimal decision rules of the DMs are given by likelihood ratio tests with constant thresholds. All equations for the general case, as well as their specialization for Gaussian statistics can be found in [12]; these are not included here due to space limitations.

We wanted to investigate the effect that different communication protocols, i.e. increasing the number of messages K, have on the performance of the team.

Empirical studies on human organizations seem to suggest that a large amount of communications is required to properly coordinate distributed DMs (there is a lot of anectodal evidence in the naval CWC doctrine which has been criticized for being communications intensive). In our study, we used an example where the observations of the DMs have Gaussian distributions with different means. In order to boost our intuition and understanding we performed numerical sensitivity analyses and studied the two limiting cases (in one DM B decides in isolation and in the other, the centralized case, DM B receives both observations) as well as two intermediate cases (the two-message and the three-message case).

3.3 Numerical Results

The baseline parameters for the Gaussian example are presented in Table 1. We can also see that as expected the three-message case is better than the two-message case, as measured by the probability of error. Notice however that the incremental improvement can be quite modest, suggesting that in this specific example one bit communication may be quite adequate. In Figure 2 we present the ROC curves of the individual DMs and in Figure 3 we compare the team ROC curves for the two and the three-message case for the organization. Notice that since the team makes a binary decision, no matter what is the number, K, of messages from A to B, the team performance can be summarized by an ROC curve. In this manner, we can directly compare the global performance of different organizations or of the same organization with different communication protocols.

In Figure 4 we see the probability of error of the team as a function of the prior probability of the null hypothesis H0 for different variances of the consulting DM and of the primary DM. The curves are symmetric and the worst performance occurs at P(H0)=0.5 since at this point prior uncertainty is maximized. These results suggest that the primary DM should be the "better" one; but this is not a general conclusion since counterexamples to this exist [11].

In Figure 5 we present the team probability of error as a function of the variance of the consulting DM and of the primary DM. In this manner, we can study the team contribution of DMs that become progressively less "capable", since we can associate increased variances of measurement noise with degraded capability of the corresponding DM (reflected for example by a degraded ROC curve.) In both cases as the variances increase the team probability of error levels off because a point is reached where the "better" DM makes the team decision alone. This is another case where the coupling between the decision rules of the two DMs is clearly demonstrated since the better DM realizes the shortcomings of the other and takes control of the team decision.

In Figure 6 we compare the centralized case, the three-message case, the two-message case and the isolation case. The improvement between the isolation case and the two-message case is 17%. The improvement between the two-message case and the three-message case was 4% and the improvement between the three-message case and the centralized case was 3%. These results suggest that for this numerical example, three messages are enough to achieve performance very close to the centralized optimal. Thus, for this team to achieve excellent results, all information does not have to be processed by a single DM. It would be nice if similar results are true for more complex decision problems, and more research is necessary along these directions.

In Figure 7 we compare the probability of error for the centralized, the three-message and the two-message cases. These results seems to re-enforce our previous conclusions since we see a bigger improvement from the two-message case to the three-message case than from the three-message case to the centralized case. Moreover, the improvement in performance from the two-message to the three-message case is greater if the primary DM is the "better" DM. Thus there is no point in giving more communication resources to a DM who is not sufficiently "smart."

3.4 Conclusions

We summarize the conclusions of the discussion above. The optimal decision rules of the two DMs are coupled and are different from the individual ones. Increased communication results to improved team performance; there is not much potential for improvement beyond the use of two bits. It is better for the team to have the smarter DM as the primary and it is not beneficial for the team to increase the communication capacity of a dumb DM. Finally, we repeat that the DMs operate as a team and make decisions in a manner that benefits the team.

4. SUMMARY OF SOME OTHER RESULTS

In the same framework, we have also examined several other problems which we shall mention only briefly because of space constraints. We have extensively worked on determining the optimal architecture of small team problems. Our results relating to teams with few DMs indicate that because of the inherent complexity of the problems no generalizations can be made; that is we need to know exactly the team members and the detailed team goal in order to determine the optimal team architecture. In general, one cannot make global statements that a particular way of organizing DMs is always better than another. There appear to be special problems in which a particular architecture is better; however, it can be inferior in other cases. These results suggest that we must abandon strict team optimality considerations if we want to obtain guidelines for designing organizations that are generally, but not always, superior.

We also examined the problem of a team of infinite identical DMs in tandem and determined necessary and sufficient conditions for the probability of error of the team to go to zero. This result, which complements the results in [8], demonstrates that it is possible for a team which consists of an infinite number of DMs to err. It also indicates that in general the architecture of a large decision-making organization should be closer to a parallel one than to a tandem one.

We also examined what we call 'recruiting' problems. In these we are given a DM and specifications which the team has to meet and we are required to find the optimal consulting DM who, when recruited to the team, will enable the team to meet predefined specifications. By optimal we mean the DM who has the smallest area under his ROC curve. We discovered that, except in some very restricted versions, these problems are extremely difficult, if not impossible, to solve. This seems to indicate that organizations should be satisfied to recruit DMs which will be performing a 'good enough' job for them and should not try to recruit the 'optimal' DM.

5. CONCLUDING REMARKS

During the past few years there has been considerable research interest in the field of decentralized hypothesis testing and significant results have been obtained.

We thus know that these type of problems are NP-complete [6]; this implies that as the number of the DMs, or of the hypotheses or of the messages increases these problems cannot be solved in any efficient manner because they become computationally intractable due to combinatorial explotion. Moreover, other results as in [11] show that even very small and restricted examples of these team problems (2 DMs, binary hypothesis and single bit communications) are also very hard to solve. Hence it has become evident that theseorganizational design problems exhibit not only great computational complexity, but also great inherent complexity.

Our original objective was to analyze and understand decentralized hypothesis testing problems, draw conclusions and then use the conclusions to achieve improved decision making while striving for global optimality. We feel that, although we have a good understanding of the difficulties associated with these problems, our conclusions thus far do not lead to improved decision making in a practical setting. Perhaps we should abandon optimality. Also, we need to develop new tools that aggregate complex organizations and represent them by simpler "equivalent" ones. Such organization aggregation methodology would be very beneficial; however, it seems that it cannot be derived in a straight-forward way. To be specific, in our paradigm of binary hypothesis testing problem, there does not appear to exist a simple computational way of deriving, say, upper and lower bounds (or other reasonable approximations) to the team ROC curve so that we can easily compare alternate organizations; see also [11]. Since real life organizations face much more complicated problems (continuous hypotheses, huge amount of data, several specializations, dynamic deadlines, limited resources etc etc.) more approximate, but computationally feasible, approaches need to be developed.

ACKNOWLEDGMENT

This research was supported by the Office of Naval Research under contract ONR/N00014-84-K-0519 (NR 649-003)

REFERENCES

[1] Van Trees,H.L.,"Detection, Estimation and Modulation Theory, vol. I", New York, Wiley, 1969

[2] Tenney, R.R., and Sandell, N.R., "Detection with Distributed Sensors," IEEE Trans. on Aerospace and Electronic Systems, AES-17, July 1981, pp.501-509

[3] Ekchian, L.K., "Optimal Design of Distributed Detection Networks," Ph.D. Dissertation, Dept. of Elec. Eng. and Computer Science, Mass. Inst. of Technology, Cambridge, Mass., 1982

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

[5] Tsitsiklis, J.N., "Problems in Decentralized Decision Making and Computation," Ph.D. Dissertation, Dept. of Elec. Eng. and Computer Science, Mass. Inst. of Technology, Cambridge, Mass., 1984

[6] Tsitsiklis J.N., and Athans, M., "On the Complexity of Decentralized and Detection Problems," IEEE Trans. on Automatic Control, 30, May 1985, pp. 440-446

[7] Tsitsiklis J.N., "Decentralized Detection by a Large Number of Sensors," Mathematics of Control, Signals and Systems, Vol.1, #2, 1988

[8] Papastavrou J.D., "Distributed Detection with Selective Communications," M.S. Dissertation, Dept. of Elec. Eng. and Computer Science, Mass. Inst. of Technology, Cambridge, Mass., 1986

[9] Reibman A.R., anb Nolte L.W., "Optimal Detection and Performance of Distributed Sensor Systems," IEEE Trans.on Aerospace and Electronic Systems,23, January 1987, pp. 24-30

[10] Reibman A.R., anb Nolte L.W., "Design and Performance Comparisons of Distributed Detection Networks," IEEE Trans.on Aerospace and Electronic Systems, 23, November 1987, pp. 789-797

[11] Papastavrou, J.D., and Athans, M., "Optimum Configuration for Distributed Teams of Two Decision Makers," Proceedings of the 1988 Symposium on Command and Control Research, 1988, pp. 162-168

[12] Pothiawala, Javed, "Analysis of a Two-Sensor Tandem Distributed Detection Network," M.S. Dissertation, Dept. of Elec. Eng. and Computer Science, Mass. Inst. of Technology, Cambridge, Mass., 1989

[13] Simon, H.A., "On How to Decide What to Do, " The Bell Journal of Economics, 1978, 9, pp. 494-507

[14] Simon, H.A., "The Sciences of the Artificial", (Second Edition), M.I.T. Press, Cambridge, Mass., 1981

[15] Athans, M., "The Expert Team of Experts to Command and Control Organizations," Proceedings of the 4th MIT/ONR Workshop, 1981

[16] Athans, M., "Command and Control Theory: A Challenge to Control Science," IEEE Trans. on Automatic Control, Vol. 32, No. 4, April 1987, pp. 286-293

[17] Abrahamson, J.A., "Launch Costs Still a Key," Aerospace America, Vol. 24, No. 4, April 1986, pp. 35-36

[18] Janis, I.L., Mann L., "Decision Making," The Free Press, New York, 1977

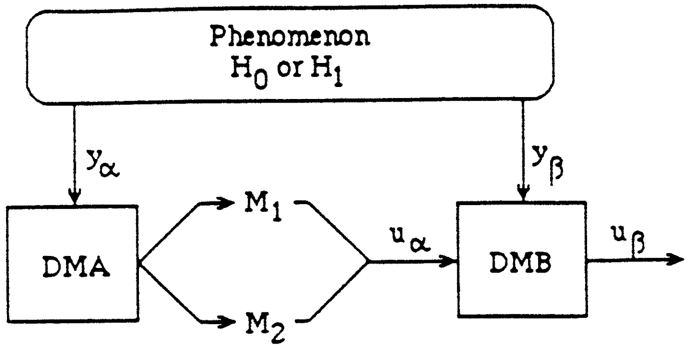
[19] Ungson, G.R., Braunstein, D.N., "Decision Making," Kent Publishing Co., Boston, 1982

Parameters used : $\mu_0 = 0$ $\mu_1 = 10$

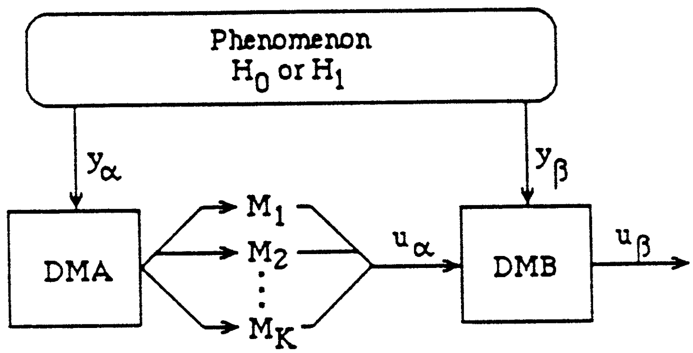
$\sigma_\alpha^2 = 100$ $\sigma_\beta^2 = 100$

$P(H_0)$	No. of messages	PR(E)
0.5	2	0.25758
	3	0.24778
0.8	2	0.16629
	3	0.16194

TABLE 1:

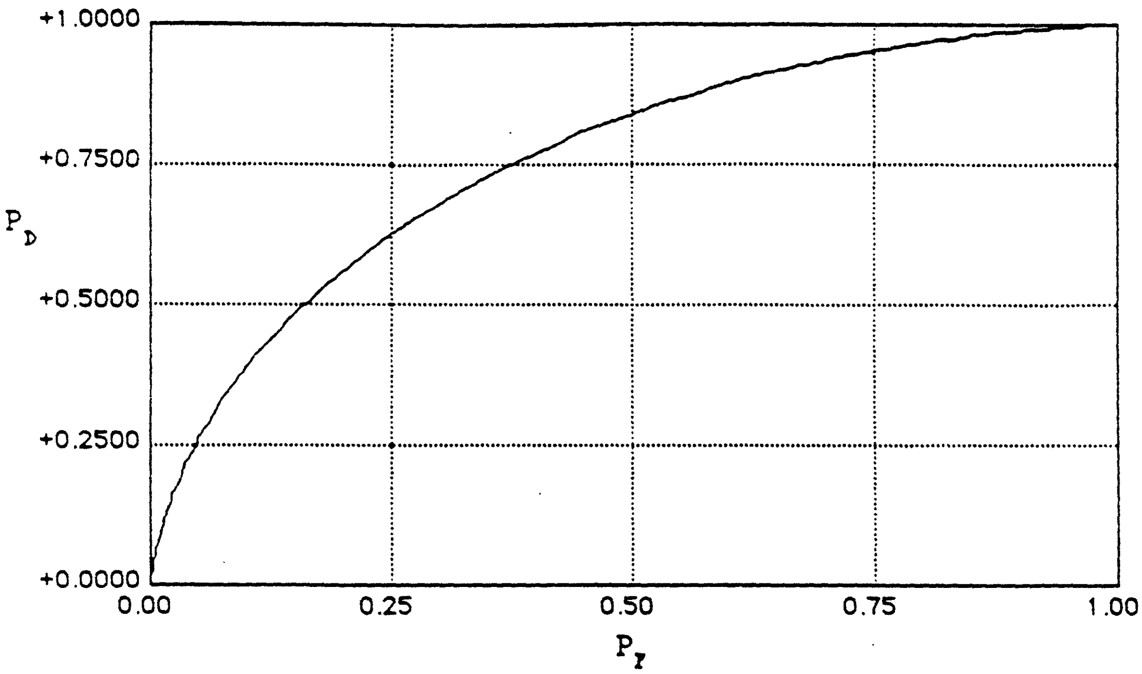


(a) Two-message (K=2) Tandem Distributed Detection Network, ($u_\alpha = \{M_1, M_2\}$)

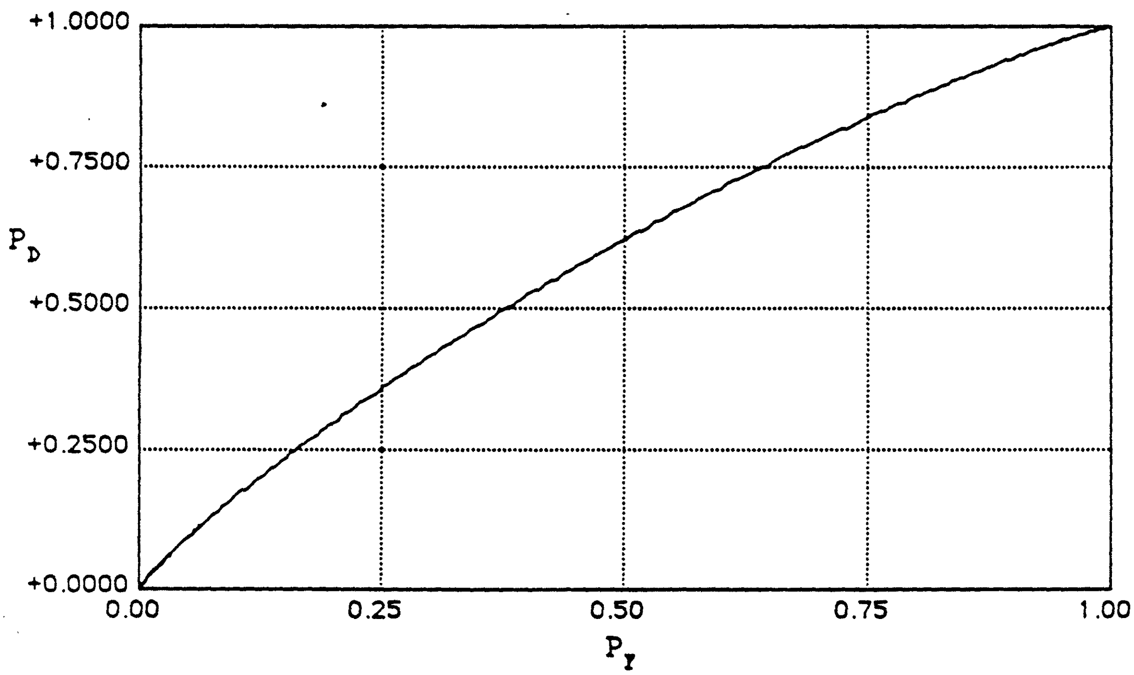


(b) K-message Tandem Distributed Detection Network, ($u_\alpha = \{M_1, M_2, ..., M_K\}$)

Figure 1: Problem Formulation

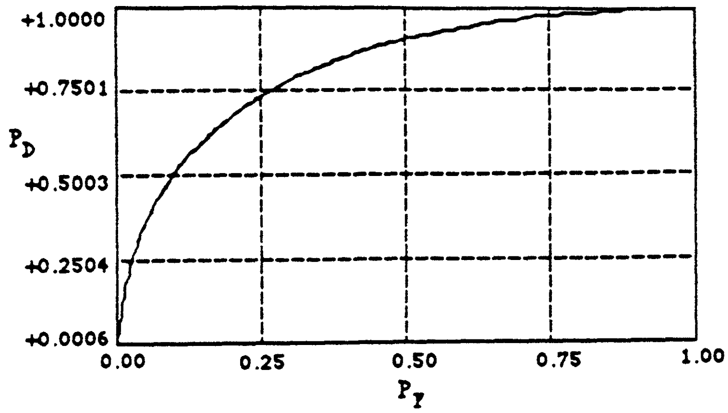


(a) ROC curve for DMA ($\sigma_\alpha^2 = 100$)

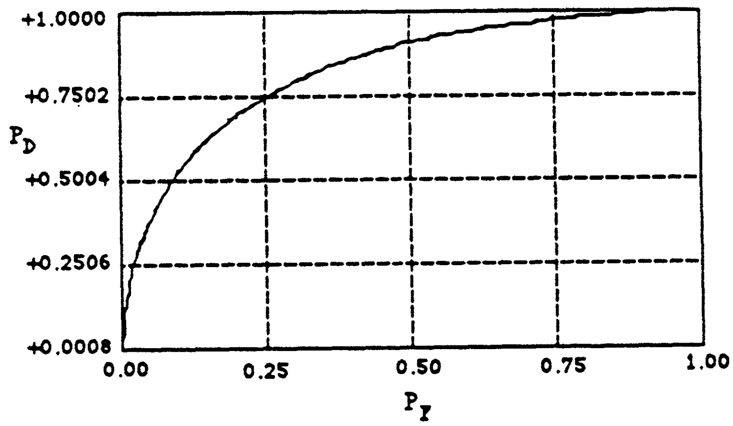


(b) ROC curve for DMB ($\sigma_\beta^2 = 1000$)

Figure 2: Individual DM's ROC Curves

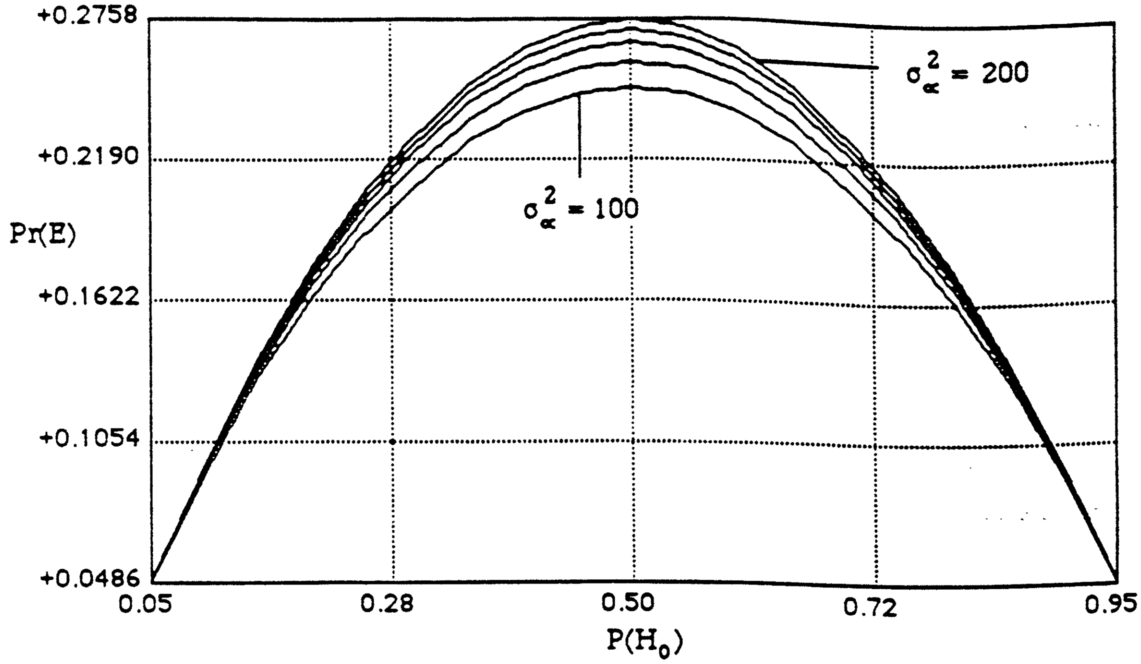


(a) Two-message Case

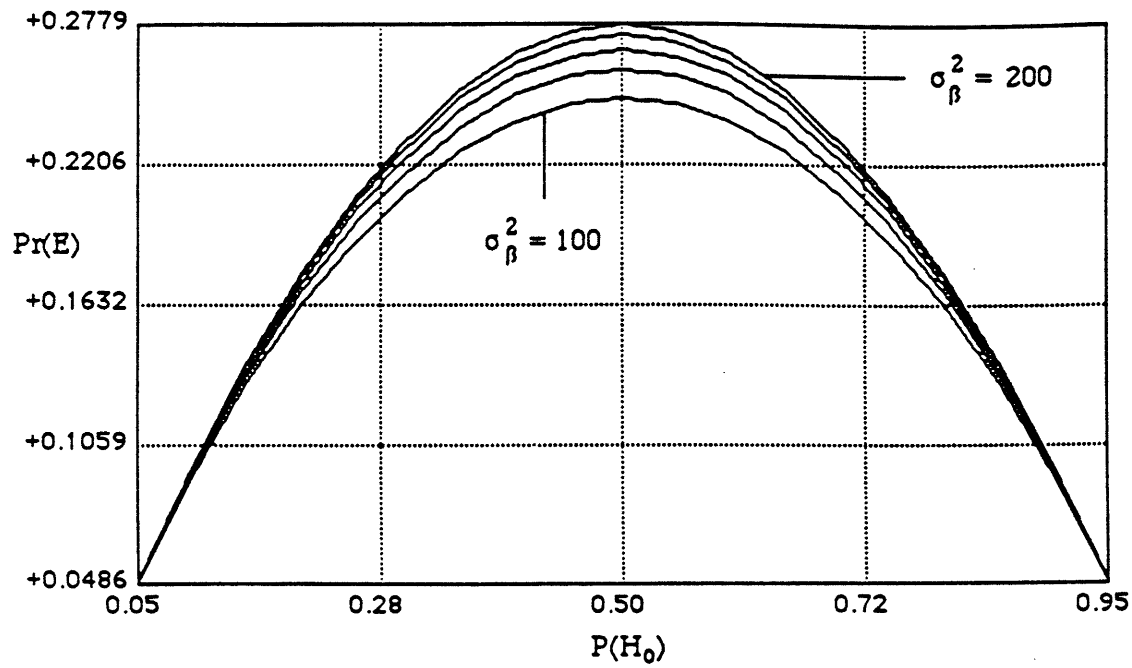


(b) Three-message Case

Figure 3: Team ROC Curves.

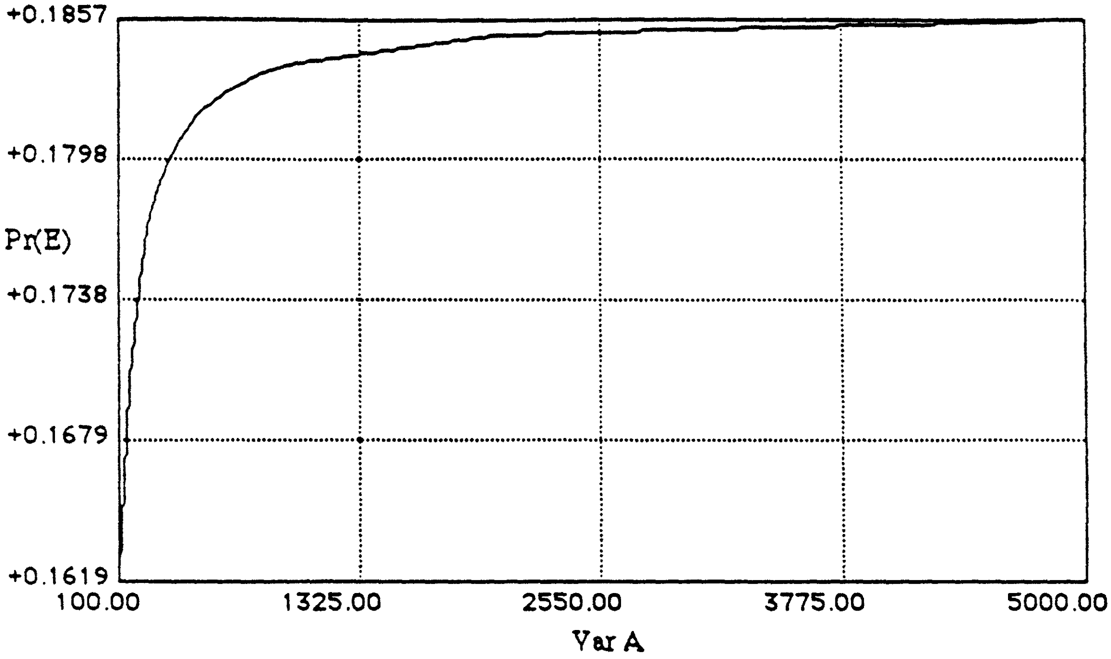


(a) $\sigma_\alpha^2 = 100, 125, 150, 175, 200$ $\sigma_\beta^2 = 100$

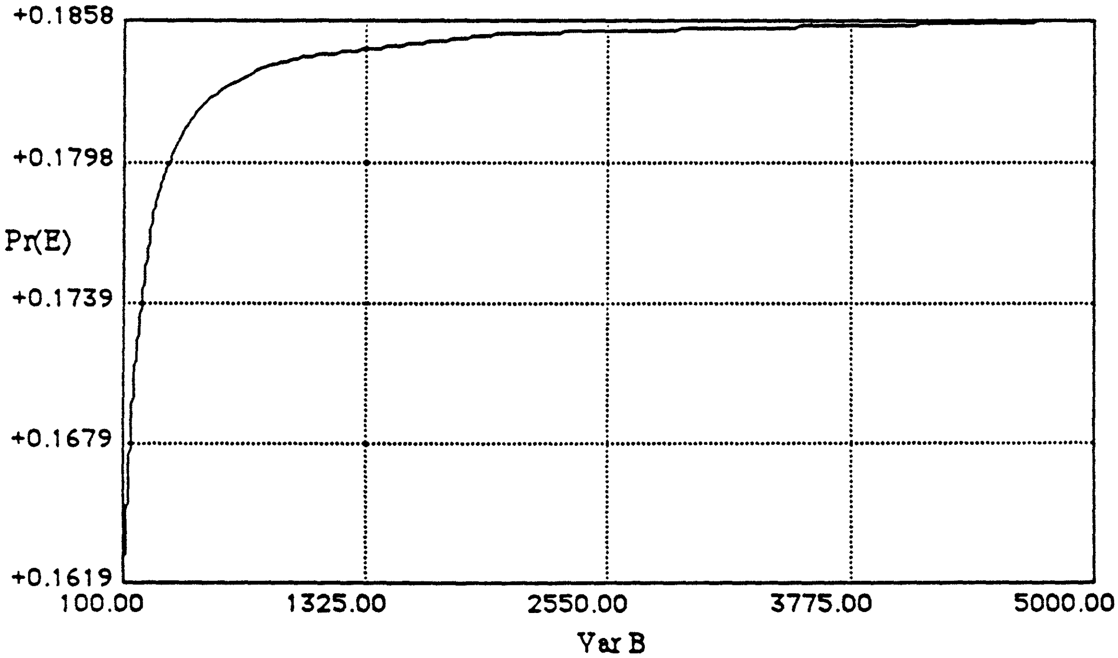


(b) $\sigma_\alpha^2 = 100$ $\sigma_\beta^2 = 100, 125, 150, 175, 200$

Figure 4: Pr(E) vs. $P(H_0)$ (K = 3)



(a) $\sigma_\beta^2 = 100$ $P(H_0) = 0.8$



$\sigma_\alpha^2 = 100$ $P(H_0) = 0.8$

Figure 5: Pr(E) vs. σ_α^2, σ_β^2 (K = 3)

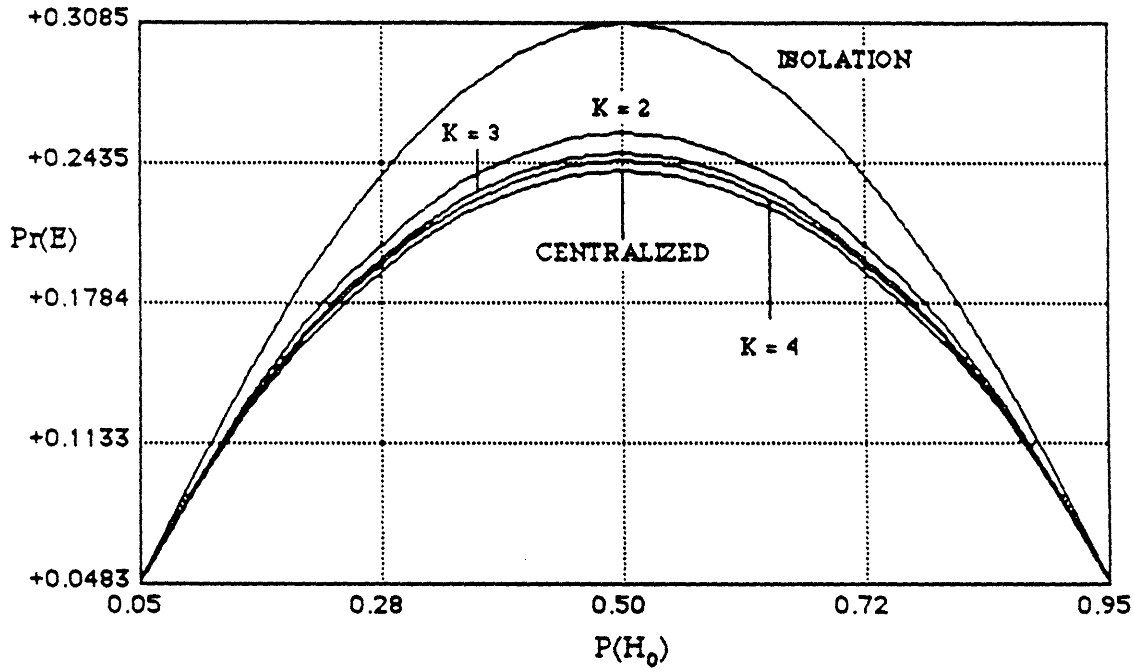
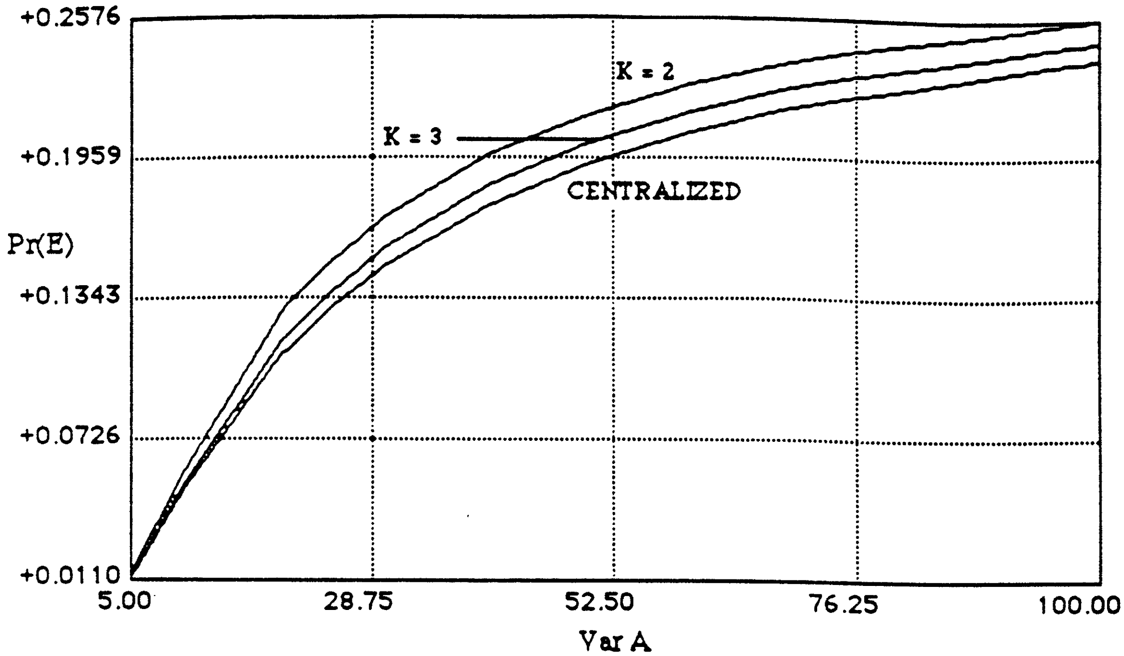
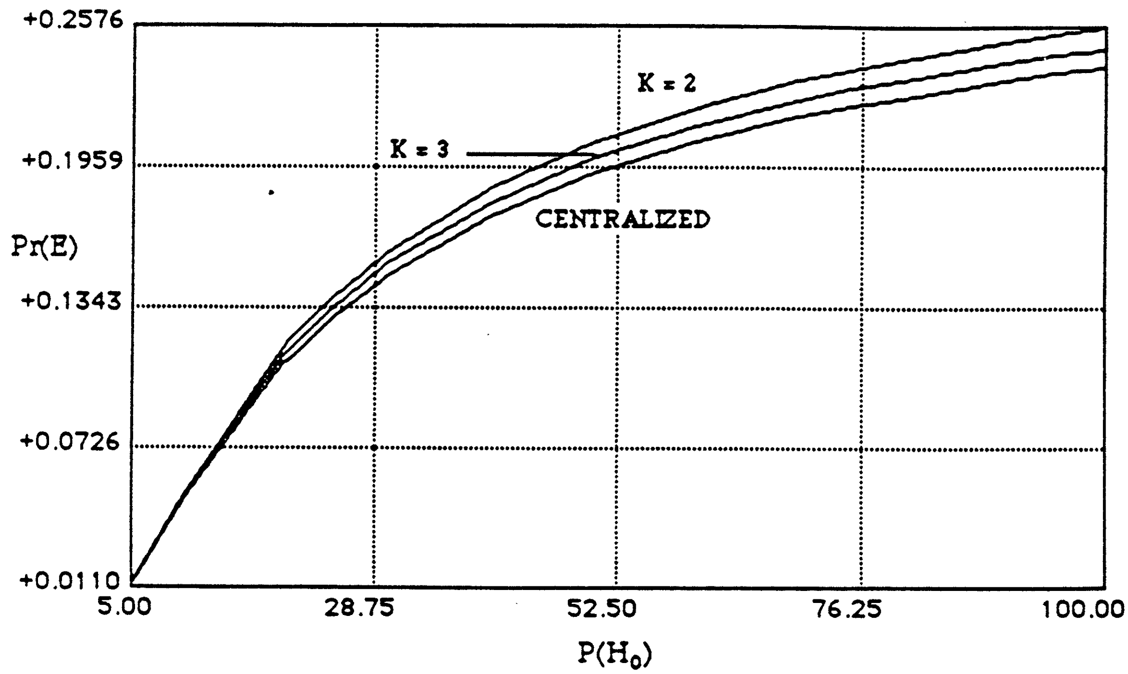


Figure 6: comparison of Performance $(\sigma_\alpha{}^2 = \sigma_\beta{}^2 = 100)$



(a) DMA smarter ($\sigma_\beta^2 = 100$)



(b) DMB smarter ($\sigma_\alpha^2 = 100$)

Figure 7: Effect of Increasing Communication