

Decentralized Quickest Change Detection

Venugopal V. Veeravalli, *Senior Member, IEEE*

Abstract—A decentralized formulation of the quickest change detection problem is studied, where the distributions of the observations at all of the sensors in the system change at the time of disruption, and the sensors communicate with a common fusion center. A Bayesian setting is considered in which *a priori* knowledge of the change time distribution is available. The observations are assumed to be independent from sensor to sensor, conditioned on the change hypothesis. An optimal solution to the problem is derived under a quasi-classical information structure, where each sensor retains only its messages from the past (restricted local memory), and receives feedback from the fusion center about the past messages of the other sensors (full feedback). A technique for implementation of the optimal solution is given, and the solution is extended to the situation where *a priori* change time distribution information is not available. The structure of the optimal solution is then used to arrive at a simple suboptimal policy that does not require any past message information. Numerical examples are given, which illustrate that the optimal solution offers little improvement over the suboptimal one, i.e., that feedback from the fusion center cannot be exploited to improve performance.

Index Terms—Bayes problem, detection, distributed decision-making, monitoring, multisensor systems, quickest change.

I. INTRODUCTION

The problem of detecting an abrupt change in a system based on stochastic observations of the system arises in a variety of applications including biomedical signal processing, quality control engineering, finance, link failure detection in communication networks, and channel monitoring for mobile wireless communication systems. The *centralized* version of this problem—where all the information about the change is available at a single location—is well-understood and has been solved under a variety of criteria since the seminal work by Page [1]. (See, e.g., [2]–[5]. For an overview of the work in this area, see [6].) However, there are situations where the information available for decision making is *decentralized*.

As an example, consider the following change detection problem in an intelligent wireless sensor network. Information about the change is available through measurements taken at several wireless sensors in the network, and a central entity (*fusion center*) must detect the change as soon as possible based on these measurements. The sensors, being lower power wireless devices, are constrained to send messages belonging to a *finite alphabet* to the fusion center. An optimal solution to this quickest change detection problem is found by a joint (*team*) optimization of all the sensor functions and the fusion center policy.

The design of quickest change detection procedures usually involves optimizing the tradeoff between two kinds of performance measures, one being a measure of detection delay and other being the a measure of the frequency of false alarms. In the centralized case, there are two standard mathematical formulations for the optimum tradeoff problem. The first of these is a minimax formulation, due to Lorden [2], in which the goal is to minimize the worst case delay subject to a lower bound

Manuscript received August 31, 1999; revised October 12, 2000. This work was supported by the Office of Naval Research under Grant N00014-97-1-0823. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Ulm, Germany, 1997.

The author is with the Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: vvv@uiuc.edu).

Communicated by U. Madhow, Associate Editor for Detection and Estimation. Publisher Item Identifier S 0018-9448(01)02849-8.

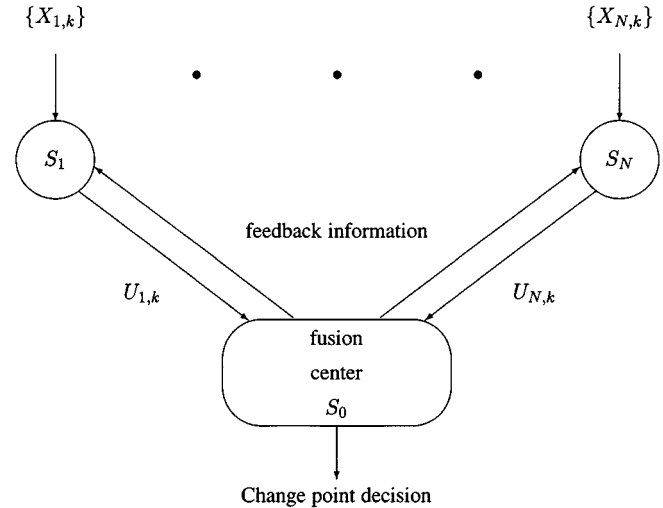


Fig. 1. General setting for decentralized change detection.

on the mean time between false alarms. The second is a Bayesian formulation, proposed by Shiryaev and Kolmogorov [7], [5] in which the change point is assumed to have a geometric prior distribution, and the goal is to minimize the expected delay subject to an upper bound on false alarm probability. For decentralized change detection, we will see that the Bayesian formulation is preferable since it allows for a dynamic programming solution. Also, to our knowledge there does not seem to be an easy way to find an optimal solution for the minimax formulation.¹ As in the centralized case, we can obtain a useful non-Bayesian test by taking a limiting form of the Bayesian solution [9].

Various sensor configurations are possible for decentralized decision making [10], [11]. We restrict our attention to the basic fusion configuration for decentralization, where a fusion center is responsible for making a final decision about the change. Consider a system with N sensors as shown in Fig. 1. At time k , an observation $X_{\ell,k}$ is made at sensor S_{ℓ} . Further, based on the information available at S_{ℓ} at time k , a message $U_{\ell,k}$, belonging to a finite alphabet of size D_{ℓ} , is formed and sent to the fusion center. We assume that two-way communication is possible between the sensors and the fusion center. In particular, at time k , the fusion center could possibly broadcast to each sensor all the sensor messages it received at time $k-1$. Based on the sequence of sensor messages, a decision about the change is made at the fusion center. It is assumed that the sensors stop taking observations as soon as the fusion center decides that the change has occurred.

Various information structures are possible for the decentralized configuration of Fig. 1 depending on how the feedback and local information is used at the sensors (see [12], [13]). We focus here on the special case of a system with full feedback and local memory restricted to past decisions, where the message $U_{\ell,k}$ formed by sensor S_{ℓ} at time k is a function of only its current observation $X_{\ell,k}$ and the past decisions of all the sensors (including itself), i.e.,

$$U_{\ell,k} = \psi_{\ell,k}(X_{\ell,k}; U_{1,[1,k-1]}, \dots, U_{N,[1,k-1]}). \quad (1)$$

This information structure is said to be *quasi-classical* [14], [15] and (as explained in [13]) appears to be the only information structure that leads to a tractable detection problem.

¹However, it is possible to find a minimax solution under restrictions on the class of sensor functions. In [8], such a solution is found for the case where the sensors simply integrate their observations and compare the result to a fixed threshold (that is optimized).

Throughout this correspondence, we assume that the observations at each sensor are independent (conditioned on the change time), have a common probability density function (pdf) before the disruption, and a different common pdf from the time of disruption. Furthermore, we assume that the observations are independent from sensor to sensor, conditioned on the disruption time. While this independence assumption is somewhat restrictive, we believe it provides a reasonable starting point for the analysis of decentralized change detection problems.

After briefly describing the problem formulation in Section II, we develop an optimal solution for the information structure considered in this correspondence in Section III. We discuss techniques that facilitate implementation of this solution in Sections IV-A and IV-B. We also discuss how we might relax the geometric prior assumption on the change time in Section IV-C. In Section V, we develop a useful *suboptimal* solution that does not require any past message information. In Section VI, we give numerical results for several illustrative examples. A surprising finding in these results is that the optimal solution offers little improvement over the considerably simpler suboptimal solution.

II. PROBLEM FORMULATION

- 1) It is assumed that the change time Γ is geometrically distributed, i.e.,

$$\mathbb{P}\{\Gamma = 0\} = \nu \quad \text{and} \quad \mathbb{P}\{\Gamma = i\} | \{\Gamma > 0\} = \rho(1 - \rho)^{i-1}.$$

The value of 0 for Γ accounts for the possibility that the disruption took place before the observations were made.

- 2) Conditioned on Γ , the observation sequences $\{X_{1,k}\}$, $\{X_{2,k}\}$, \dots , $\{X_{N,k}\}$ are assumed to be mutually independent. Furthermore, it is assumed that the observations in a particular sequence, say $\{X_{\ell,k}\}$, are independent conditioned on Γ , have a common pdf $f_0^{(\ell)}$ before the disruption, and common pdf $f_1^{(\ell)}$ from the time of disruption.
- 3) The sensor messages $U_{\ell,k} \in \{0, 1, \dots, D_{\ell} - 1\}$. Past sensor message information at time k is denoted by I_{k-1} and is given by

$$I_{k-1} = \{U_{1,[1,k-1]}, U_{2,[1,k-1]}, \dots, U_{N,[1,k-1]}\}.$$

- 4) The quasi-classical information structure, specified in (1), is assumed. Thus,

$$U_{\ell,k} = \psi_{\ell,k}(X_{\ell,k}; I_{k-1}).$$

Note that $\psi_{\ell,k}$ can be regarded as a *quantizer* of the observation $X_{\ell,k}$ that depends on I_{k-1} , i.e.,

$$U_{\ell,k} = \psi_{\ell,k}(X_{\ell,k}; I_{k-1}) \equiv \phi_{\ell,k}(I_{k-1})(X_{\ell,k}). \quad (2)$$

In our analysis, we will generally drop the “ (I_{k-1}) ” to keep the notation from getting cumbersome, and it is implicitly assumed that $\phi_{\ell,k}$ depends on I_{k-1} .

- 5) The fusion center policy χ consists of selecting a stopping time τ at which time it is decided that the disruption has occurred.
- 6) For notational convenience, bold-faced symbols are used to denote collections of variables and functions across sensors, e.g., $\mathbf{U}_k = (U_{1,k}, \dots, U_{N,k})$, $\phi_k = (\phi_{1,k}, \dots, \phi_{N,k})$, etc.

To each choice of fusion center policy and sensor quantizer functions, there correspond two types of performance indexes. The first is the probability of *false alarm*

$$P_{\text{FA}} = \mathbb{P}\{\tau < \Gamma\}$$

and the second is the expected detection delay

$$E_{\text{DD}} = \mathbb{E}\{(\tau - \Gamma)^+\}$$

where $x^+ = \max\{0, x\}$. The design problem at hand is one of optimizing the tradeoff between these two performance indexes.

There are two ways to pose the optimum tradeoff problem [5].

1) *Variational Formulation*: Minimize E_{DD} over all admissible choices of fusion center policy χ and sensor quantizers $\{\phi_k\}_{k=1}^{\infty}$, subject to $P_{\text{FA}} \leq \alpha$, where α is a control parameter.

2) *Bayesian Formulation*: Define the total Bayes cost (risk)

$$R(c) = P_{\text{FA}} + cE_{\text{DD}} = \mathbb{E}[\mathbb{I}_{\{\tau < \Gamma\}} + c(\tau - \Gamma)^+] \quad (3)$$

where \mathbb{I} denotes the indicator function, and the constant $c > 0$ may be interpreted as the cost of each unit of delay. The Bayesian optimization problem can then be stated as

$$\text{minimize } R(c) \text{ over all admissible choices of } \chi \text{ and } \{\phi_k\}_{k=1}^{\infty}. \quad (4)$$

The following alternative, and useful, expression for the Bayes risk of (3) can be derived using arguments similar to those given in [5, pp. 151–152]

$$R(c) = \mathbb{P}\{\Gamma > \tau\} + c \mathbb{E} \left[\sum_{k=1}^{\tau-1} \mathbb{P}\{\Gamma \leq k\} \right]. \quad (5)$$

Also, as in the centralized version of the change detection problem, the solution to the variational problem is easily obtained once the Bayes solution is found. In particular, we have the following result whose proof is nearly identical to the corresponding proof in the centralized case [5].

Theorem 1: An optimal policy for the variational formulation is a Bayes policy for an appropriately chosen value of the tradeoff parameter c .

Our solution to (4) is based on dynamic programming (DP) [16] arguments. The logical steps involved are similar to those found in [12], where a decentralized sequential detection problem is solved. This is to be expected since the quickest change detection and sequential detection problems are closely related [17].

The general solution given in the following section is somewhat cumbersome; we refer the reader to a tutorial article on sequential decision fusion problems [13], where the solution is presented for the special case of two sensors sending one bit at each time step to the fusion center ($N = 2$, $D = 2$).

III. BAYESIAN SOLUTION FOR THE QUASI-CLASSICAL INFORMATION STRUCTURE

In order to address the solution to the Bayes problem of (4), we first restrict the stopping time τ to a finite horizon, say the interval $[0, T]$ (we will remove this restriction in Section III-C). Since the Bayes risk $R(c)$ of (5) is additive over time, minimization over the finite horizon can be done recursively using the following DP approach.

Since the decision about the change is made at the fusion center, the (minimum) expected cost-to-go at time k is a function of the information available for decision making at the fusion center at time k , i.e., I_k . We denote this cost-to-go by $\tilde{J}_k^T(I_k)$.

Based on (5) it is easy to show (see [16, p. 133] for a similar example) that

$$\tilde{J}_T^T(I_T) = \mathbb{P}\{\Gamma > T\} | I_T \quad (6)$$

and for $0 \leq k \leq T$

$$\bar{J}_k^T(I_k) = \min \left\{ P(\{\Gamma > k\} | I_k), cP(\{\Gamma \leq k\} | I_k) + \min_{\phi_{k+1}} \mathbb{E} \left[\bar{J}_{k+1}^T(I_{k+1}) | I_k \right] \right\} \quad (7)$$

with the understanding that I_0 is the empty set. The first term in the outer minimum is the cost of stopping at time k and deciding that a change has taken place, and the second term is the cost continuing at time k . Note that the minimum expected cost for the finite horizon optimization problem is simply \bar{J}_0^T .

A. The Structure of Optimal Sensor Quantizers

We will now establish that optimal sensor quantizers can be found within a structured class of functions admitting a finite-dimensional parameterization. This is the class of likelihood ratio quantizers defined below.

Definition 1: Consider a sensor observation X with pdf f_1 and f_0 , under H_1 and H_0 , respectively.

- a) A function ϕ , that maps X to the finite alphabet $\{0, 1, \dots, D-1\}$, is called a *monotone likelihood ratio quantizer* (MLRQ) if there exist thresholds $\lambda_1, \dots, \lambda_{D-1}$ satisfying

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{D-1} \leq \infty$$

such that

$$\phi(x) = d \quad \text{only if } \lambda_d < L(x) \leq \lambda_{d+1}, \quad d = 0, \dots, D-1,$$

where $\lambda_0 = 0, \lambda_D = \infty$, and $L(x) = f_1(x)/f_0(x)$ is the *likelihood ratio* between f_1 and f_0

- b) A sensor quantizer ϕ is called a *likelihood ratio quantizer* (LRQ) if there exists a permutation mapping

$$\Sigma: \{0, \dots, D\} \mapsto \{0, \dots, D\}$$

such that the composite function $\Sigma \circ \phi$ is an MLRQ.

Note that in the special case of binary sensor messages ($D = 2$), the LRQs in the above definition reduce to the standard binary likelihood ratio tests.

The LRQ structure of optimal sensor quantizers is established in the following theorem whose proof is given in the Appendix.

Theorem 2: Optimal finite horizon sensor quantizers can be found in the class of LRQs with thresholds that depend on the past decision information.

In the special case of $D_\ell = 2$, optimal sensor quantizers are described by

$$U_{\ell,k}^* = \phi_{\ell,k}^*(X_{\ell,k}) = \begin{cases} 1, & \text{if } L_\ell(X_{\ell,k}) > \lambda_k^{(\ell)}(I_{k-1}) \\ 0, & \text{otherwise} \end{cases}$$

where $\lambda_k^{(\ell)} \equiv \lambda_{1,k}^{(\ell)}$. In general, the LRQ at time k at sensor S_ℓ is described in terms of $D_\ell - 1$ thresholds as in Definition 1. We will see later in Section III-B1) that optimal sensors quantizers actually belong to the smaller set of MLRQs.

B. Sufficient Statistic for Dynamic Programming

In the following, we explore the DP solution of (6) and (7) further and show that they can be rewritten in terms of a one-dimensional sufficient statistic p_k defined by

$$p_k = P(\{\Gamma \leq k\} | I_k). \quad (8)$$

We begin by obtaining a recursive equation for p_k

$$\begin{aligned} p_{k+1} &= P(\{\Gamma \leq k+1\} | I_{k+1}) \\ &= P(\{\Gamma \leq k+1\} | I_k, \mathbf{U}_{k+1}) \\ &= \frac{P(\{\Gamma \leq k+1\} | I_k) f(\mathbf{U}_{k+1} | I_k; \Gamma \leq k+1)}{f(\mathbf{U}_{k+1} | I_k)} \end{aligned} \quad (9)$$

where f above is used to denote the probability mass function (pmf) of \mathbf{U}_{k+1} . Now, the terms in the numerator and denominator of (9) can be expanded as

$$\begin{aligned} P(\{\Gamma \leq k+1\} | I_k) &= P(\{\Gamma \leq k\} | I_k) + P(\{\Gamma = k+1\} | I_k) \\ &= p_k + P(\{\Gamma = k+1\} | I_k; \{\Gamma \geq k+1\}) P(\{\Gamma \geq k+1\} | I_k) \\ &= p_k + \rho(1-p_k) \end{aligned} \quad (10)$$

and

$$f(\mathbf{U}_{k+1} | \{\Gamma \leq k+1\}; I_k) = \prod_{\ell=1}^N P_{f_1^{(\ell)}} \{ \phi_{\ell, k+1}(X_{\ell, k+1}) = U_{\ell, k+1} \} \quad (11)$$

and

$$\begin{aligned} f(\mathbf{U}_{k+1} | I_k) &= f(\mathbf{U}_{k+1} | I_k; \{\Gamma \leq k+1\}) P(\{\Gamma \leq k+1\} | I_k) \\ &\quad + f(\mathbf{U}_{k+1} | I_k; \{\Gamma > k+1\}) P(\{\Gamma > k+1\} | I_k). \end{aligned} \quad (12)$$

Let X_ℓ denote a generic observation at S_ℓ , and let $\phi_\ell(x_\ell) \in \{0, 1, \dots, D_\ell - 1\}$. Now define the functions g and h as follows:

$$g(\mathbf{d}; \phi; p) = [p + (1-p)\rho] \prod_{\ell=1}^N P_{f_1^{(\ell)}} \{ \phi_\ell(X_\ell) = d_\ell \} \quad (13)$$

$$\begin{aligned} h(\mathbf{d}; \phi; p) &= g(\mathbf{d}; \phi; p) + (1-p)(1-\rho) \\ &\quad \cdot \prod_{\ell=1}^N P_{f_0^{(\ell)}} \{ \phi_\ell(X_\ell) = d_\ell \} \end{aligned} \quad (14)$$

where $\mathbf{d} = (d_1, \dots, d_N)$ with $d_\ell \in \{0, 1, \dots, D_\ell - 1\}$.

From (9)–(14), it is clear that the following equation holds:

$$p_{k+1} = \frac{g(\mathbf{U}_{k+1}; \phi_{k+1}; p_k)}{h(\mathbf{U}_{k+1}; \phi_{k+1}; p_k)} \quad (15)$$

with $p_0 = P(\Gamma = 0) = \nu$.

Note that (15) is not necessarily a recursive equation, since ϕ_{k+1} depends on I_k in general. However, we will establish in the following theorem that optimal ϕ_{k+1} depend on I_k only through p_k . Thus, (15) is indeed a recursion for p_k if optimal sensor quantizers are used.

Theorem 3:

- i) For each $k, 0 \leq k \leq T$, the function $\bar{J}_k^T(I_k)$ can be written as a function of only p_k , say $\bar{J}_k^T(p_k)$.
- ii) For each $k, 0 \leq k \leq T-1$, an optimal sensor quantizer at time $k+1$ depends on I_k only through p_k .

Proof: The proof is identical to that of [12, proof of Proposition 3]. \square

Based on this result, the finite-horizon DP equations can be rewritten in terms of p_k as follows:

$$J_T^T(p_T) = 1 - p_T \quad (16)$$

and for $k = 0, \dots, T-1$

$$J_k^T(p_k) = \min \left\{ (1-p_k), c p_k + A_k^T(p_k) \right\} \quad (17)$$

where

$$A_k^T(p) = \min_{\phi} \sum_{\mathbf{d}} J_{k+1}^T \left(\frac{g(\mathbf{d}; \phi; p)}{h(\mathbf{d}; \phi; p)} \right) h(\mathbf{d}; \phi; p),$$

for all $p \in [0, 1]$. (18)

Some useful properties of the functions J_k^T and A_k^T are given in the following lemma which is easily proven by simple induction arguments using (16)–(18).

Lemma 1: The functions $J_k^T(p)$ and $A_k^T(p)$ are nonnegative, concave functions of p , for $p \in [0, 1]$. Furthermore, $A_k^T(1) = J_k^T(1) = 0$, $\forall k$.

1) *Optimal Finite-Horizon Quantizers:* We established earlier in this section that optimal finite-horizon sensor quantizers are LRQs with thresholds that depend on the past decision information. Theorem 3 further implies that the thresholds of the optimal LRQs depend on the I_k only through p_k . Now, a simple permutation argument can be used to argue that optimal sensor quantizers at time $k+1$ can be found in the smaller class of MLRQs with thresholds that depend on p_k . Thus, if we define the set Φ_M to be the set of MLRQs, then $A_k^T(p)$ of (18) can be written as

$$A_k^T(p) = \min_{\phi \in \Phi_M} \sum_{\mathbf{d}} J_{k+1}^T \left(\frac{g(\mathbf{d}; \phi; p)}{h(\mathbf{d}; \phi; p)} \right) h(\mathbf{d}; \phi; p),$$

for all $p \in [0, 1]$. (19)

Finally, since an MLRQ $\tilde{\phi}_{\ell, k+1}$ is completely characterized by $D_\ell - 1$ thresholds, the minimization to obtain $A_k^T(p_k)$ in (19) reduces to a finite-dimensional optimization problem over $\sum_{l=1}^N (D_\ell - 1)$ thresholds.

C. Infinite-Horizon Optimization

As derived in [12], it is easy to establish that

$$\lim_{T \rightarrow \infty} J_k^T(p) = \inf_{T: T > k} J_k^T(p) = J_k^\infty(p) = J(p)$$

where the last equality follows from the fact that the geometric distribution of Γ is memoryless.

Taking limits as $T \rightarrow \infty$ in (17) and (18), we get that the infinite horizon cost-to-go function $J(p)$ satisfies the Bellman equation

$$J(p) = \min\{(1-p), cp + A_J(p)\} \quad (20)$$

where

$$\begin{aligned} A_J(p) &= \lim_{T \rightarrow \infty} A_k^T(p) \\ &= \min_{\phi \in \Phi_M} \sum_{\mathbf{d}} J \left(\frac{g(\mathbf{d}; \phi; p)}{h(\mathbf{d}; \phi; p)} \right) h(\mathbf{d}; \phi; p), \quad \text{for all } p \in [0, 1]. \end{aligned} \quad (21)$$

Note that the minimum Bayes risk is simply $J(p)$. In addition, if $J(p)$ is computed for all p , then the optimal policy of the fusion center can be obtained easily from the right-hand side of (20). However, it is possible to obtain the qualitative structure of the optimal fusion center policy without actually computing $J(p)$.

1) *The Structure of the Optimal Fusion Center Policy:* We begin with a lemma which follows straightforwardly by taking limits as $T \rightarrow \infty$ in Lemma 1.

Lemma 2: The functions $J(p)$ and $A_J(p)$ are nonnegative concave functions of p , for $p \in [0, 1]$. Furthermore, they satisfy the endpoint condition

$$A_J(1) = J(1) = 0.$$

Using (20) and Lemma 2, it is easy to show that an optimal fusion center policy will have the threshold structure given in the following theorem.

Theorem 4: An optimal fusion center policy has stopping time τ that is given by

$$\tau = \inf\{k: p_k > a\} \quad (22)$$

where a is the unique solution to

$$ca + A_J(a) = 1 - a$$

when the solution exists. If the curves $cp + A_J(p)$ and $(1-p)$ do not intersect on $[0, 1]$, a is taken to be 0.

Remark 1: If $a = 0$ and $\nu > 0$, the decision of the fusion center is that a disruption took place before the observations were taken.

While Theorem 4 completes the solution to the Bayesian optimization problem of (4), the preceding analysis does not specify a clear “recipe” for computing the thresholds for the optimal sensor quantizers and the fusion center. In the following, we give a key result that facilitates the computation of these thresholds.

2) *Uniqueness of $J(p)$ and its Consequences:* Let $\mathcal{S} \subset C[0, 1]$ be the set of all concave functions on $[0, 1]$ that are bounded (in sup norm) by the function $(1-p)$, $p \in [0, 1]$. It is clear that the infinite-horizon cost-to-go function J belongs to the set \mathcal{S} . For any $G \in \mathcal{S}$, we define

$$W_G(\phi; p) := \sum_{\mathbf{d}} G \left(\frac{g(\mathbf{d}; \phi; p)}{h(\mathbf{d}; \phi; p)} \right) h(\mathbf{d}; \phi; p).$$

Clearly, $A_J(p)$ of (21) can be written in terms of W_G as

$$A_J(p) = \min_{\phi \in \Phi_M} W_J(\phi; p).$$

Now, let the mapping $\mathcal{T}: \mathcal{S} \rightarrow \mathcal{S}$ be defined by

$$\mathcal{T}G(p) = \min \left\{ (1-p), cp + \min_{\phi \in \Phi_M} W_G(\phi; p) \right\}, \quad \text{for } G \in \mathcal{S}. \quad (23)$$

From (20), it is clear that J is a fixed point of \mathcal{T} . The following result is that this fixed point is *unique*. The proof is nearly identical to that of [12, Theorem 3], and is hence omitted.

Theorem 5: The infinite-horizon cost-to-go function J is the unique fixed point of \mathcal{T} .

An important consequence of this result is that $J(p)$ can be obtained by successive approximation. Indeed, we can show using a simple induction argument that $T^{n+1}\eta(p) \leq T^n\eta(p)$, for each $p \in [0, 1]$, where $\eta(p) = 1-p$. This means that $T^n\eta$ converges monotonically to J as $n \rightarrow \infty$. Theorem 5 also implies that a *stationary* set of sensor quantizers is optimal (also see [12]).

IV. NUMERICAL COMPUTATION AND IMPLEMENTATION

In the following, we exploit the above two consequences of Theorem 5 to arrive at a recipe for numerical computation of the optimal solution.

A. Threshold Computation

Note that the stationary set of optimal quantizers are MLRQs. Let the thresholds for the MLRQ at sensor S_ℓ be denoted by $\lambda_i^{(\ell)}$, $i = 1, 2, \dots, D_\ell - 1$. Furthermore, we collect the threshold values at all

the sensors in the vector λ . Since the MLRQs are completely characterized by the thresholds λ , we may rewrite the functions g and h of (13) and (14) as

$$g(\mathbf{d}; \lambda; p) = [p + (1-p)\rho] \prod_{\ell=1}^N \mathbf{P}_{f_1^{(\ell)}} \left\{ \lambda_{d_{\ell-1}}^{(\ell)} < L_{\ell}(X_{\ell}) \leq \lambda_{d_{\ell}}^{(\ell)} \right\} \quad (24)$$

$$h(\mathbf{d}; \lambda; p) = g(\mathbf{d}; \lambda; p) + (1-p)(1-\rho) \prod_{\ell=1}^N \mathbf{P}_{f_0^{(\ell)}} \left\{ \lambda_{d_{\ell-1}}^{(\ell)} < L_{\ell}(X_{\ell}) \leq \lambda_{d_{\ell}}^{(\ell)} \right\} \quad (25)$$

where X_{ℓ} denotes a generic observation at S_{ℓ} , and $\lambda_0^{(\ell)} = 0$, $\lambda_{D_{\ell}}^{(\ell)} = \infty$, for all ℓ .

Using the definitions of g and h given in (24) and (25), we can rewrite the mapping \mathcal{T} of (23) as

$$\mathcal{T} G(p) = \min \left\{ (1-p), cp + \min_{\lambda} W_G(\lambda, p) \right\} \quad (26)$$

where, for $G \in \mathcal{S}$,

$$W_G(\lambda, p) = \sum_a G \left(\frac{g(\mathbf{d}; \lambda; p)}{h(\mathbf{d}; \lambda; p)} \right) h(\mathbf{d}; \lambda; p). \quad (27)$$

To obtain J , we successively apply \mathcal{T} to $\eta(p) = 1 - p$. This can be done numerically with p appropriately quantized on $[0, 1]$. Since convergence to J is guaranteed by the results of Section III-C2), we can stop the iteration as soon as the norm difference between successive iterates falls below a desired threshold. The optimal sensor thresholds (as functions of p) are, of course, obtained from the last iteration as

$$\lambda_{\text{opt}}(p) = \arg \min_{\lambda} W_J(\lambda, p) \quad (28)$$

and the optimum fusion threshold a is obtained by solving (numerically)

$$(1-a) = ca + W_J(\lambda_{\text{opt}}(p), p). \quad (29)$$

B. Implementation of the Optimal Solution

To implement the optimal solution, both the sensors and the fusion center need to compute the sufficient statistic p_k recursively. There is obviously a one time-step delay between the recursions at the sensors versus that at the fusion center. Using the new definitions of g and h given in (24) and (25), we can rewrite the recursion (15) as

$$p_{k+1} = \frac{g(\mathbf{U}_{k+1}; \lambda_{\text{opt}}(p); p_k)}{h(\mathbf{U}_{k+1}; \lambda_{\text{opt}}(p); p_k)}, \quad \text{with } p_0 = \nu \quad (30)$$

where $\lambda_{\text{opt}}(p)$ is as specified in (28). Implementation of this recursion is facilitated by making the variable substitution

$$q_k = \frac{p_k}{\rho(1-p_k)}. \quad (31)$$

In particular, it is easy to show that

$$q_{k+1} = \frac{q_k + 1}{1-\rho} \prod_{\ell=1}^N L_{\ell}^U(U_{\ell, k+1}, q_k), \quad \text{with } q_0 = \frac{\nu}{\rho(1-\nu)} \quad (32)$$

where L_{ℓ}^U is the likelihood ratio of the pmf induced on sensor decisions through the sensor quantizer, i.e.,

$$L_{\ell}^U(i, q) = \frac{\mathbf{P}_{f_1^{(\ell)}} \left\{ \hat{\lambda}_{i-1}^{(\ell)}(q) < L_{\ell}(X_{\ell}) \leq \hat{\lambda}_i^{(\ell)}(q) \right\}}{\mathbf{P}_{f_0^{(\ell)}} \left\{ \hat{\lambda}_{i-1}^{(\ell)}(q) < L_{\ell}(X_{\ell}) \leq \hat{\lambda}_i^{(\ell)}(q) \right\}} \quad (33)$$

where

$$\hat{\lambda}_i^{(\ell)}(q) = \lambda_{\text{opt}, i}^{(\ell)} \left(\frac{\rho q}{1 + \rho q} \right). \quad (34)$$

Also, as before, X_{ℓ} denotes a generic observation at S_{ℓ} , and $\hat{\lambda}_0^{(\ell)}(q) = 0$, $\hat{\lambda}_{D_{\ell}}^{(\ell)}(q) = \infty$, for all ℓ .

The sensor messages at time k are formed through MLRQs based on q_{k-1} , i.e.,

$$U_{\ell, k} = d_{\ell}, \quad \text{only if } \hat{\lambda}_{d_{\ell}}^{(\ell)}(q_{k-1}) < L_{\ell}(X_{\ell, k}) \leq \hat{\lambda}_{d_{\ell+1}}^{(\ell)}(q_{k-1}).$$

Finally, the fusion center stopping time is given by

$$\tau = \inf \left\{ k: q_k > \frac{a}{\rho(1-a)} \right\}.$$

C. Relaxing the Geometric Prior Assumption

The Bayesian analysis of the preceding sections depended heavily on the assumption that the prior distribution of the change time was geometric. For the *centralized* change detection problem, it is known that in the absence of prior information, a useful test may be obtained from a limiting form of the Bayes solution, by letting ρ and ν go to zero. This limiting form was first suggested by Girshick and Rubin [9], and is sometimes referred to as the GRS (Girshick, Rubin, and Shirayev) procedure.

For the decentralized change detection problem, there does not seem to be an obvious way to find the optimal solution (for both the sensor and fusion center decision rules) under Lorden's minimax criterion. As mentioned in Section I, a minimax solution was found in [8], only under a restriction on the class on sensor functions. The Bayesian solution we have found can, of course, be applied to the situation where prior information is not available, in a similar fashion as in the centralized case. In particular, if we take limits in (32) as $\rho, \nu \rightarrow 0$, in such a way that $q_0 \rightarrow 0$, then we get the following recursion for q_k :

$$q_{k+1} = [q_k + 1] \prod_{\ell=1}^N L_{\ell}^U(U_{\ell, k+1}, q_k), \quad \text{with } q_0 = 0.$$

Also, the MLRQ thresholds at the sensors become independent of q_k . In particular, from (34), we get

$$\hat{\lambda}_i^{(\ell)}(q) = \lambda_{\text{opt}, i}^{(\ell)}(0), \quad \text{for all } q.$$

Finally, the fusion center stopping rule is as in the centralized GRS procedure, i.e.,

$$\tau = \inf \{k: q_k > A\}$$

where A is varied to the tradeoff E_{DD} versus P_{FA} .

V. CONSTANT-THRESHOLD SUBOPTIMAL POLICY

It is of interest to compare the performance of the optimal solution with that of a suboptimal policy that ignores all the past message infor-

mation and simply uses MLRQs with constant thresholds at the sensors, i.e.,

$$U_{\ell, k} = d_\ell \quad \text{only if } \tilde{\lambda}_{d_\ell}^{(\ell)} < L_\ell(X_{\ell, k}) \leq \tilde{\lambda}_{d_\ell+1}^{(\ell)}.$$

The question then arises as to how these constant thresholds $\tilde{\lambda}$ should be chosen. With the sensor thresholds fixed at $\tilde{\lambda}$, the fusion center faces a centralized change detection problem with independent and identically distributed (i.i.d.) vector “observations” \mathbf{U}_k . The pmf induced on $U_{\ell, k}$ by the constant-threshold quantizer at S_ℓ is given by

$$f_{j, \ell}^U(i) = \mathbf{P}_{f_j^{(\ell)}} \left\{ \tilde{\lambda}_{i-1}^{(\ell)} < L_\ell(X_\ell) \leq \tilde{\lambda}_i^{(\ell)} \right\}, \\ i = 0, \dots, D_\ell - 1, \quad j = 0, 1. \quad (35)$$

As we shall see in the numerical results given in the next section, for the centralized change detection problem, the slope of the tradeoff curve between $\ln \text{P}_{\text{FA}}$ and E_{DD} is roughly equal to the negative of the Kullback–Leibler (KL) “distance” between the distributions after and before the change. Hence, we should expect that the best tradeoff between P_{FA} and E_{DD} is obtained when $D(f_{1, \ell}^U, f_{0, \ell}^U)$, the KL distance between $f_{1, \ell}^U$ and $f_{0, \ell}^U$, is maximum. Thus, a reasonable choice for $\tilde{\lambda}^{(\ell)}$ is the one that maximizes $D(f_{1, \ell}^U, f_{0, \ell}^U)$.

The optimum fusion center solution (for fixed $\tilde{\lambda}$) is obtained by computing the sufficient statistic \tilde{q}_k recursively as

$$\tilde{q}_{k+1} = \frac{\tilde{q}_k + 1}{1 - \rho} \prod_{\ell=1}^N \tilde{L}_\ell^U(U_{\ell, k+1}), \quad \text{with } \tilde{q}_0 = \frac{\nu}{\rho(1 - \nu)}$$

where

$$\tilde{L}_\ell^U(i) = \frac{f_{1, \ell}^U(i)}{f_{0, \ell}^U(i)}.$$

The fusion center stopping time is then given by

$$\hat{\tau} = \inf \left\{ k: \tilde{q}_k > \frac{a}{\rho(1 - a)} \right\}.$$

VI. NUMERICAL RESULTS

We illustrate the theoretical results of the previous sections through two simple examples.

Example 1: Assume that the sensor observations are Poisson random variables with different means before and after the disruption. Let the observations at sensor S_ℓ have mean $\mu_{0\ell}$ before the disruption, and mean $\mu_{1\ell}$ after the disruption. Without loss of generality assume that $\mu_{1\ell} > \mu_{0\ell}$. Then the likelihood ratio at S_ℓ is given by

$$L_\ell(x) = \left(\frac{\mu_{1\ell}}{\mu_{0\ell}} \right)^x \exp \{ -(\mu_{1\ell} - \mu_{0\ell}) \}.$$

Note that the likelihood ratio is monotonically increasing in x .

Example 2: Assume that the sensor observations are Gaussian random variables with different means, but the same variance, before and after the disruption. Again, let the observations at sensor S_ℓ have mean $\mu_{0\ell}$ before the disruption, and mean $\mu_{1\ell}$ after the disruption, with $\mu_{1\ell} > \mu_{0\ell}$. Then the likelihood ratio at S_ℓ is given by

$$L_\ell(x) = \exp \left\{ \frac{(\mu_{1\ell} - \mu_{0\ell})x}{\sigma_\ell^2} - \frac{\mu_{1\ell}^2 - \mu_{0\ell}^2}{2\sigma_\ell^2} \right\}$$

where σ_ℓ^2 is the variance. Note that $L_\ell(x)$ is monotonically increasing in x in this example as well.

Since L_ℓ is monotonically increasing, we can characterize the sensor quantizers in terms of thresholds on the observations, rather than on their likelihood ratios. To further simplify the examples, we assume that the sensor messages are binary, i.e., $D_\ell = 2$ for all ℓ . Then the sensor

quantizers reduce to binary likelihood ratio tests that are characterized by a single threshold, i.e.,

$$U_{\ell, k} = \begin{cases} 1, & \text{if } X_{\ell, k} \geq \lambda^{(\ell)}(p_{k-1}) \\ 0, & \text{if } X_{\ell, k} < \lambda^{(\ell)}(p_{k-1}). \end{cases}$$

Furthermore, the functions g and h of (24) and (25) needed for the DP solution simplify to

$$g(\mathbf{d}; \boldsymbol{\lambda}; p) = [p + (1-p)\rho] \prod_{\ell=1}^N \left[P_1 \left\{ X_\ell > \lambda^{(\ell)} \right\} \right]^{d_\ell} \\ \cdot \left[P_1 \left\{ X_\ell \leq \lambda^{(\ell)} \right\} \right]^{1-d_\ell} \\ h(\mathbf{d}; \boldsymbol{\lambda}; p) = g(\mathbf{d}; \boldsymbol{\lambda}; p) + (1-p)(1-\rho) \prod_{\ell=1}^N \left[P_0 \left\{ X_\ell > \lambda^{(\ell)} \right\} \right]^{d_\ell} \\ \cdot \left[P_0 \left\{ X_\ell \leq \lambda^{(\ell)} \right\} \right]^{1-d_\ell}.$$

Also, the equation for L_ℓ^U (see (33)) which is required in the recursions for q_k becomes

$$L_\ell^U(i, q) = \frac{\left[P_1 \left\{ X_\ell > \hat{\lambda}^{(\ell)}(q) \right\} \right]^i \left[P_1 \left\{ X_\ell \leq \hat{\lambda}^{(\ell)}(q) \right\} \right]^{1-i}}{\left[P_0 \left\{ X_\ell > \hat{\lambda}^{(\ell)}(q) \right\} \right]^i \left[P_0 \left\{ X_\ell \leq \hat{\lambda}^{(\ell)}(q) \right\} \right]^{1-i}}$$

where $\hat{\lambda}^{(\ell)}(q) = \lambda_{\text{opt}}^{(\ell)}(\rho q / (1 + \rho q))$.

In optimizing the thresholds, we first consider the two sensor case ($N = 2$). Optimal sensor threshold functions ($\lambda_{\text{opt}}^{(1)}, \lambda_{\text{opt}}^{(2)}$) are obtained by minimizing $W_G(\lambda^{(1)}, \lambda^{(2)}, p)$ over $(\lambda^{(1)}, \lambda^{(2)}) \in \mathbb{R}^2$ at each stage of the successive approximation procedure.

As a further simplification, we consider the symmetric situation where the sensors have identical statistics. Through numerical experimentation for a few test cases, it was found that the optimum thresholds were identical (functions of p) at the two sensors² for both examples. We hence simplify the optimization by setting $\lambda^{(1)} = \lambda^{(2)} = \lambda$ and optimizing W_G over the single threshold λ . Sample results for the Poisson and Gaussian examples are shown in Figs. 2 and 3. The optimum fusion center threshold a is easily found from $J(p)$ as shown in the figures.

It is interesting to note that the optimal sensor threshold functions are nonmonotonic. We also note that $\lambda_{\text{opt}}^{(\ell)}(p)$ is discontinuous³ for the Gaussian example. The general trend seems to be to progressively favor the change hypothesis as p increases toward a ; however, there are abrupt “corrections” to this trend.

Note that Bayes cost corresponding to the optimal solution is given by $J(\nu)$. For the examples shown in Figs. 2 and 3, $\nu = 0$, and hence $J(0)$ is the minimum possible value of $\text{P}_{\text{FA}} + c\text{E}_{\text{DD}}$, for the given value of c . However, the individual performance metrics P_{FA} and E_{DD} cannot be obtained from $J(0)$. These quantities may be obtained, for each value of c , via Monte Carlo simulations. The pairs of $(\text{P}_{\text{FA}}, \text{E}_{\text{DD}})$ values thus obtained form a *tradeoff curve* for the optimal policy.

We may compare different policies by plotting their tradeoff curves on the same plot. The three policies of interest are i) the optimal centralized policy; ii) the optimal solution; and iii) the suboptimal solution of Section V. The tradeoff curves for all three policies are obtained via Monte Carlo simulations for various values of c . The tradeoff curve for the optimal centralized policy is obtained by varying the single-threshold parameter at the fusion center. The tradeoff curves

²It is known that in decentralized detection problems, assuming identical distributions at the sensors does not necessarily result in identical solutions for the optimal sensor quantizers, except asymptotically as $N \rightarrow \infty$ [10].

³As noted in [12], this behavior might be surprising at first, but such behavior is commonly observed in control systems where “bang–bang” control is optimal.

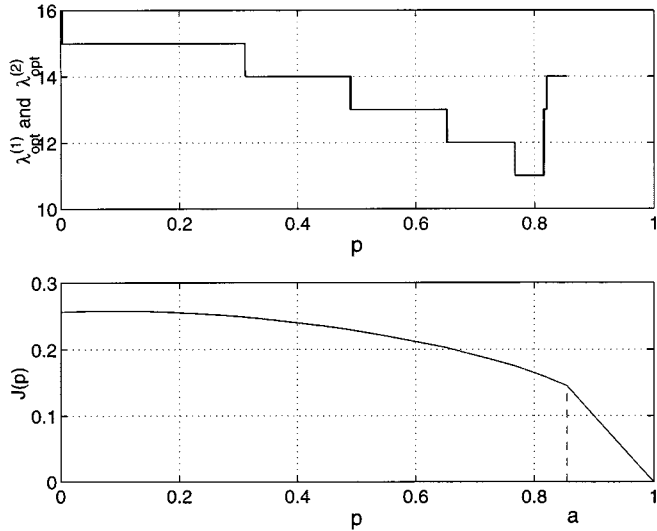


Fig. 2. Optimum thresholds for two identical sensors sending binary messages—Poisson example. Sensor observations are Poisson with mean 10 before the disruption, and mean 14 after the disruption. The parameter values are $c = 0.06$, $\rho = 0.05$, and $\nu = 0$. Results were obtained using 200 iterations of the successive approximation procedure with 5000 points on the p axis. The optimal sensor threshold function is only shown for $p \leq a$, since it is not used for $p > a$. The fusion center threshold a is marked on the plot showing $J(p)$.

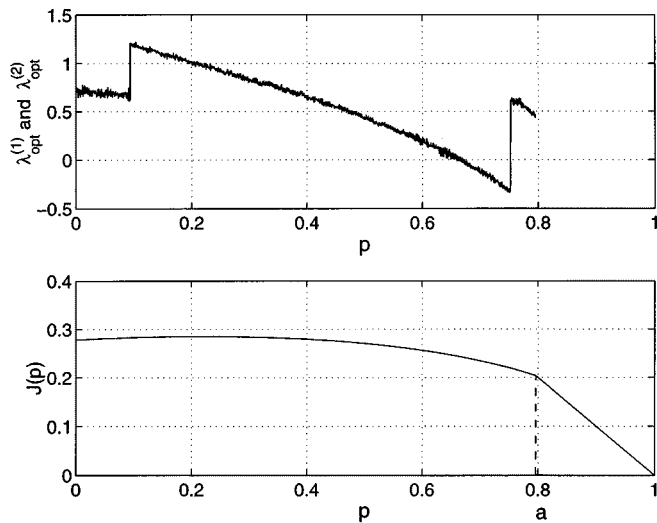


Fig. 3. Optimum thresholds for two identical sensors sending binary messages—Gaussian example. Sensor observations are $\mathcal{N}(0, 1)$ and $\mathcal{N}(0.75, 1)$ before and after the disruption, respectively. The parameter values are $c = 0.1$, $\rho = 0.3$, and $\nu = 0$. Results were obtained using 200 iterations of the successive approximation procedure with 5000 points on the p axis. The sensor threshold was quantized to 500 values in the range from $[-1, 1.75]$.

for the Poisson and Gaussian examples with two identical sensors are shown in Figs. 4 and 5, respectively. Also shown in Fig. 6 are tradeoff curves corresponding to two sensors with different Poisson statistics. Similar results are obtained for systems with five sensors.

It is interesting to see that for the centralized policy, the plot of $\ln P_{FA}$ and E_{DD} is a straight line with slope that is approximately equal to

$$-\sum_{\ell} D(f_1^{(\ell)}, f_0^{(\ell)}).$$

For the suboptimal policy, the tradeoff curve between $\ln P_{FA}$ and E_{DD} has slope that is roughly equal to

$$-\sum_{\ell} D(f_{1,\ell}^U, f_{0,\ell}^U)$$

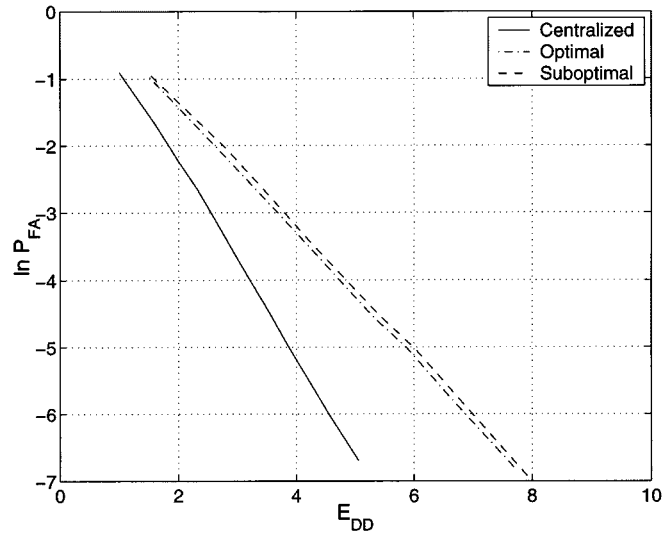


Fig. 4. Tradeoff curves for two identical Poisson sensors. Sensor observations are Poisson with mean 10 before the disruption, and mean 14 after the disruption. The parameter values are $\rho = 0.05$ and $\nu = 0$. The KL distance for the sensor observations is 0.7106. The threshold that maximizes the KL distance at the output of the sensor is $\tilde{\lambda} = 13$, and the corresponding maximum KL distance is 0.4473.

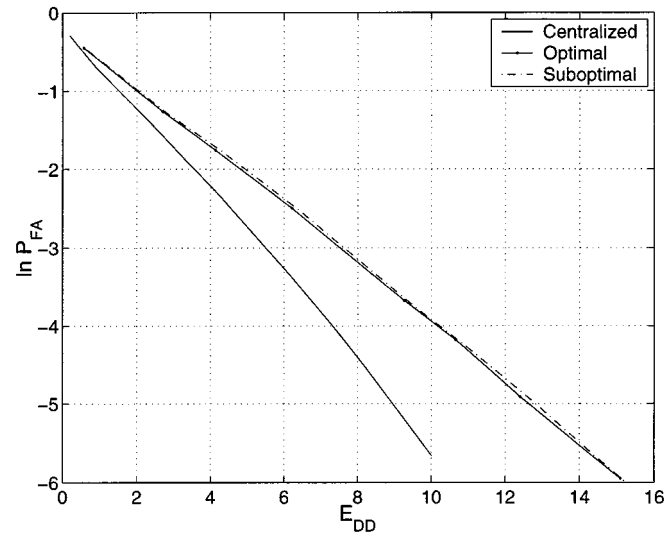


Fig. 5. Tradeoff curves for two identical Gaussian sensors. Sensor observations are $\mathcal{N}(0, 1)$ and $\mathcal{N}(0.75, 1)$ before and after the disruption, respectively. The parameter values are $\rho = 0.05$ and $\nu = 0$. The KL distance for the sensor observations is 0.28125. The threshold that maximizes the KL distance at the output of the sensor is $\tilde{\lambda} = 0.62$, and the corresponding maximum KL distance is 0.1791.

where $f_{j,\ell}^U$ is as defined in (35) corresponding to the KL distance-maximizing threshold.

The surprising observation in the performance comparison plots is that in all cases, the suboptimal policy and the optimal policy have roughly the same performance. These results indicate that a simple, constant threshold quantizer of the observations may work just as well as the optimal policy. In other words, each sensor may ignore all the past message information without loss of optimality.

VII. CONCLUSION

We studied a decentralized extension of the quickest change detection problem and showed that the problem was tractable under a quasi-classical information structure. In particular, we have given an

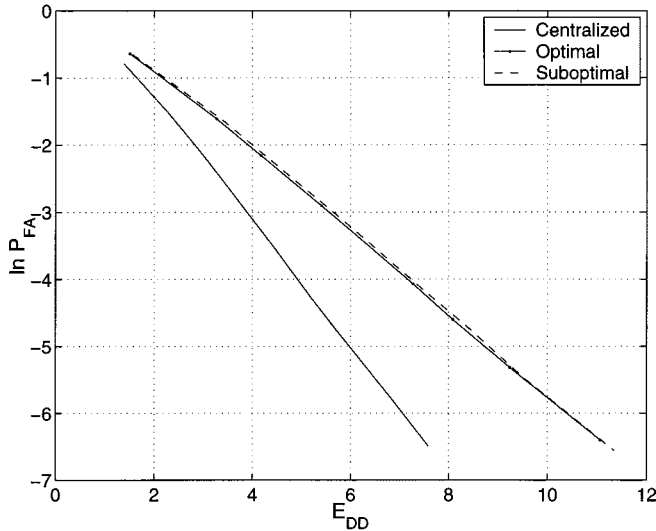


Fig. 6. Tradeoff curves for two sensors with different Poisson statistics. Observations at both sensors are Poisson with mean 10 before the disruption. The means after the disruption are 12 and 14 for sensors S_1 and S_2 , respectively. The parameter values are $\rho = 0.05$ and $\nu = 0$. The KL distance of the observations are 0.1879 and 0.7106 at sensors S_1 and S_2 , respectively. The threshold that maximizes the KL distance at the output of the sensor S_1 is $\bar{\lambda} = 11$, and the corresponding maximum KL distance is 0.119. For sensor S_2 , the values are 13 and 0.4473, as in Fig. 4.

explicit computational technique for an optimal solution for this case. We have also suggested a simplified implementation for this optimal solution. It is of interest to develop a better understanding of the non-monotonic, discontinuous behavior of the optimum sensor threshold functions. It might be possible, for example, to find simple parametric models for these functions that would facilitate implementation of the optimal solution.

Surprisingly, however, the numerical results indicate that it may not be necessary to go through the trouble of computing (and implementing) the optimal solution. In particular, using a simple constant threshold in place of the rather complicated optimal sensor threshold function does not seem to result in any performance loss in the examples we considered. We believe that the near optimality of the constant threshold policy may stem from our assumption that the sensor observations are independent, conditioned on the change point, but we have not been able to establish a rigorous result in this direction. It is also of interest to study systems in which the sensor observations are dependent, conditioned on the change point, to see if past message information is more useful in this scenario.

APPENDIX

Proof of Theorem 2: From (7), by replacing $k + 1$ by k in the second minimum, it follows that optimal ϕ_k minimize

$$R_k = \mathbb{E} \left[\tilde{J}_k^T(I_k) \mid I_{k-1} \right]$$

where the expectation is taken with respect to the joint distribution of the random variables $X_{1,k}, X_{2,k}, \dots, X_{N,k}$ conditioned on I_{k-1} . We will establish all person-by-person optimal (p.b.p.o.) solutions ϕ_k , and hence the globally optimal solution, have the LRQ structure. To this end, we fix all the quantizers in the set ϕ_k except $\phi_{\ell,k}$. To minimize R_k with respect to $\phi_{\ell,k}$ we write R_k as

$$R_k = \mathbb{E} \left[\tilde{J}_k^T(U_{1,k}, \dots, U_{N,k}, I_{k-1}) \mid I_{k-1} \right] \\ = \mathbb{E} \left[\mathbb{E} \left[\tilde{J}_k^T(U_{1,k}, \dots, U_{N,k}, I_{k-1}) \mid I_{k-1}; \Gamma \right] \mid I_{k-1} \right]$$

where the inner expectation is with respect to (w.r.t.) the joint distribution of $X_{1,k}, \dots, X_{\ell-1,k}, X_{\ell+1,k}, \dots, X_{N,k}$, conditioned on

Γ and I_{k-1} , and the outer expectation is w.r.t. the joint distribution of $X_{\ell,k}$ and Γ , conditioned on I_{k-1} . The conditioning w.r.t. $X_{\ell,k}$ is dropped in the inner expectation due to the independence of the sensor observations given Γ . The result of the inner expectation is a function of I_{k-1} , Γ , and $U_{\ell,k}$, say $K(U_{\ell,k}, I_{k-1}; \Gamma)$. Therefore,

$$R_k = \mathbb{E} [K(U_{\ell,k}, I_{k-1}; \Gamma) \mid I_{k-1}] \\ = \mathbb{E} \left[\sum_{\gamma=0}^{\infty} K(U_{\ell,k}, I_{k-1}; \gamma) \mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}; X_{\ell,k}) \mid I_{k-1} \right].$$

Now, for $\gamma \leq k$, an application of Bayes rule gives

$$\mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}; X_{\ell,k}) = \frac{\mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}) f_1^{(\ell)}(X_{\ell,k})}{f_I(X_{\ell,k} \mid I_{k-1})}$$

where $f_I(\cdot \mid I_{k-1})$ denotes conditional pdf of $X_{\ell,k}$ given I_{k-1} . Similarly, for $\gamma > k$

$$\mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}; X_{\ell,k}) = \frac{\mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}) f_0^{(\ell)}(X_{\ell,k})}{f_I(X_{\ell,k} \mid I_{k-1})}.$$

Hence,

$$R_k = \mathbb{E} \left[\frac{1}{f_I(X_{\ell,k} \mid I_{k-1})} \left(f_1^{(\ell)}(X_{\ell,k}) \sum_{\gamma=0}^k K(U_{\ell,k}, I_{k-1}; \gamma) \right. \right. \\ \cdot \mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}) + f_0^{(\ell)}(X_{\ell,k}) \\ \left. \left. \cdot \sum_{\gamma=k+1}^{\infty} K(U_{\ell,k}, I_{k-1}; \gamma) \mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}) \right) \mid I_{k-1} \right].$$

Minimizing R_k with respect to $\phi_{\ell,k}$ is equivalent to minimizing the quantity inside the above expectation almost everywhere. Thus, every p.b.p.o. solution for the quantizer at sensor S_{ℓ} at time k satisfies

$$U_{\ell,k}^* = \phi_{\ell,k}^*(X_{\ell,k}) \\ = \arg \min_{d_{\ell} \in \{0, \dots, D_{\ell}-1\}} \\ \cdot \left[H_1(d_{\ell}, I_{k-1}) f_1^{(\ell)}(X_{\ell,k}) + H_0(d_{\ell}, I_{k-1}) f_0^{(\ell)}(X_{\ell,k}) \right]$$

where

$$H_1(d_{\ell}, I_{k-1}) = \sum_{\gamma=0}^k K(d_{\ell}, I_{k-1}, \gamma) \mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1})$$

and

$$H_0(d_{\ell}, I_{k-1}) = \sum_{\gamma=k+1}^{\infty} K(d_{\ell}, I_{k-1}, \gamma) \mathbb{P}(\{\Gamma = \gamma\} \mid I_{k-1}).$$

Thus,

$$\phi_{\ell,k}^*(X_{\ell,k}) \\ = \begin{cases} \arg \min_{d_{\ell}} H_1(d_{\ell}, I_{k-1}) L_{\ell}(X_{\ell,k}) \\ \quad + H_0(d_{\ell}, I_{k-1}), & \text{if } L_{\ell}(X_{\ell,k}) < \infty \\ \arg \min_{d_{\ell}} H_1(d_{\ell}, I_{k-1}), & \text{if } L_{\ell}(X_{\ell,k}) = \infty. \end{cases} \quad (\text{A1})$$

From (A1), it is not difficult to see that $\phi_{\ell,k}^*$ is an LRQ with thresholds that depend on I_{k-1} , since it is written as a minimum of linear functions of the likelihood ratio. This completes the proof. \square

REFERENCES

- [1] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.
- [2] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [3] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.
- [4] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, no. 1, pp. 206–227, 1985.
- [5] A. N. Shiryaev, *Optimal Stopping Rules*. New York: Springer-Verlag, 1978.
- [6] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Englewood Cliffs: Prentice-Hall, 1993.

- [7] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Prob. and Applic.*, vol. 8, pp. 22–46, 1963.
- [8] R. W. Crow and S. C. Schwartz, "Quickest detection for sequential decentralized decision systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 267–283, Jan. 1996.
- [9] M. A. Girshick and H. Rubin, "A Bayes approach to a quality control model," *Ann. Math. Statist.*, vol. 23, pp. 114–125, 1952.
- [10] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI Press, 1993, vol. 2.
- [11] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1996.
- [12] V. V. Veeravalli, T. Başar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, pp. 433–442, Mar. 1993.
- [13] V. V. Veeravalli, "Sequential decision fusion: Theory and applications," *J. Franklin Inst.*, vol. 336, pp. 301–322, Feb. 1999.
- [14] Y. C. Ho, "Team decision theory and information structures," *Proc. IEEE*, vol. 68, pp. 644–654, June 1980.
- [15] T. Başar and J. B. Cruz, "Concepts and methods in multi-person coordination and control," in *Optimization and Control of Dynamic Operational Research Models*, S. G. Tzafestas, Ed. Amsterdam, The Netherlands: North-Holland, 1982, pp. 351–394.
- [16] D. Bertsekas, *Dynamic Programming*. Upper Saddle River, NJ: Prentice-Hall, 1987.
- [17] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag, 1985.

On the Optimality of Finite-Level Quantizations for Distributed Signal Detection

Jun Hu and Rick S. Blum, *Member, IEEE*

Abstract—Distributed multiple sensor detection problems with quantized observations are investigated for cases of nonbinary hypothesis and possibly statistically dependent observations from sensor to sensor conditioned on the hypothesis. The observations available at each sensor are quantized to produce a multiple digit sensor decision which is sent to a fusion center. At the fusion center, the sensor decisions are combined to form a final decision using a predetermined fusion rule. First, it is demonstrated that there is a maximum number of digits that should be used to communicate the sensor decision from a given sensor to the fusion center. This maximum is based on the number of digits used to communicate the decisions from all the other sensors to the fusion center. If more than this maximum number of digits is used, the performance of the optimum scheme will not be improved. In some special cases of great interest, the upper bound on the number of digits that should be used can be made significantly smaller. Secondly, the optimum way to allocate a fixed overall number of digits across sensors is investigated. Illustrative numerical results are also presented in this correspondence.

Index Terms—Distributed signal detection, multiple digit decisions, nonbinary hypothesis, quantizations.

I. INTRODUCTION

Signal detection algorithms which process quantized observations taken from multiple sensors continue to attract attention [1]–[5]. Such

Manuscript received March 1, 1999; revised August 1, 2000. This work was supported by the Office of Naval Research under Grant N00014-97-1-0774 and by the National Science Foundation under Grant MIP-9703730.

The authors are with the Electrical Engineering and Computer Science Department, Lehigh University, Bethlehem, PA 18015 USA (e-mail: juh2@eecs.lehigh.edu; rblum@eecs.lehigh.edu).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(01)02846-2.

algorithms have been classified as distributed signal detection algorithms. The majority of distributed signal detection research has focused on cases with statistically independent observations, binary hypothesis testing problems, and binary sensor decisions [6], [7]. Research on nonbinary hypothesis testing problems are relatively rare. An early paper on this topic [8] provided equations describing the necessary conditions for the optimum sensor processing. A more complete discussion which includes a thorough treatment of the necessary conditions for the case of independent observations is given in [9]. A nice discussion of the complexity of cases with dependent observations is also given in [9]. Neither [8] nor [9] give any numerical examples. A numerical procedure for finding the optimum processing scheme was provided in [10] for cases with dependent observations and nonbinary hypothesis and a few numerical examples are provided. However, studies of the properties of optimum schemes have been lacking. In this correspondence, we demonstrate that no more than a certain number of digits should be used to communicate a sensor decision from a particular sensor to the fusion center. Using more than that number of digits at a given sensor is unnecessary and will not lead to better performance. The number of digits which should be used is related to the number of digits used to communicate the other sensor decisions to the fusion center.

The following is a simple example of distributed detection: two sensors, binary hypothesis and additive Gaussian noise. Assume the observations at these two sensors are

$$\begin{aligned} \mathbf{H}_0: y_1 &= -1 + w_1, & y_2 &= -1 + w_2 \\ \mathbf{H}_1: y_1 &= 1 + w_1, & y_2 &= 1 + w_2 \end{aligned}$$

where w_1 and w_2 are zero-mean independent Gaussian random variables with variances 3 and 2, respectively. We also assume hypotheses \mathbf{H}_0 and \mathbf{H}_1 have equal prior probabilities.

For a centralized detection system, the optimum decision rule, the maximum *a posteriori* (MAP) rule, is relatively easy to find. On the other hand, finding an optimum distributed detection system is much more complicated. Generally, we need to try all possible fusion rules and use numerical techniques to determine the corresponding optimum decision rules.

In the following, we provide optimum decision rules without discussing how they are found. This will be discussed later in the correspondence. Here we denote a distributed scenario where the first sensor makes n_1 decisions and the second sensor makes n_2 decisions by (n_1, n_2) . The resulting optimum decision regions and P_e for the centralized system and the distributed scenarios (1,1), (1,2), and (1,3) are illustrated in Fig. 1. Of course, the centralized system, shown in Fig. 1(a), should achieve the lowest probability of error. For distributed systems, at first the more bits used at the second sensor, the lower the probability of error achieved, as shown by Fig. 1(b) and (c). However, when $n_2 > 2$, extra decision bits at the second sensor cannot decrease the probability of error further. This is illustrated by Fig. 1(c) and (d) since, in either case, $P_e = 0.1974$ although the system in Fig. 1(d) uses an extra bit.

The results for the distributed scenarios (0,4), (1,3), (2,2), and (3,1) are provided in Fig. 2. From Fig. 2, we observe that the (2,2) scenario yields the lowest probability of error in these four systems. The interesting fact here is that the signal-to-noise ratios (SNRs) at these two sensors are different. Their observations are not equally reliable.

Both these observations occur under a general set of conditions to be given in this correspondence. The material is organized as follows. Section II presents the model for the observations and the distributed decision making. In Section III, we show that there is a maximum number of digits that should be used for the sensor decision at a given