Quickest Change Detection in Distributed Sensor Systems

Alexander G. Tartakovsky *
Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, CA 90089-1113, USA
tartakov@math.usc.edu

Venugopal V. Veeravalli †
Dept of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
vvv@uiuc.edu

Abstract – In the conventional formulation of the change-point detection problem, there is a sequence of observations whose distribution changes at some unknown point in time, and the goal is to detect this change as quickly as possible, subject to false alarm constraints. It is known that in the case where the observations are independent and identically distributed (iid) and the change point is modeled as a geometrically distributed random variable, the Shiryaev detection procedure minimizes the expected detection delay, subject to a constraint on the false alarm probability. In this paper, we present effective decentralized detection procedures for the multi-sensor situation where the information available for decision-making is distributed across a set of sensors. We present asymptotically optimal procedures for two scenarios. In the first scenario, the sensors send quantized versions of their observations to a fusion center where the change detection is performed based on all the sensor messages. In the second scenario, the sensors perform local change detection using Shiryaev-Roberts procedures and send their final decisions to the fusion center for combining. We show that our decentralized procedures for latter scenario have the same first order asymptotic performance as the centralized Shiryaev-Roberts procedure that has access to all of the sensor observations. We also present numerical results for a simple example involving Gaussian observations.

Keywords: Change-point problems, quickest detection, sequential detection, distributed decisions, optimal fusion, multi-sensor, Shiryaev procedure, Shiryaev-Roberts procedure.

1 Introduction

An important application area for distributed sensor systems is environment surveillance and monitoring. Specific applications include: intrusion detection in computer networks, intrusion detection in security systems, chemical or biological warfare agent detection systems to protect against terrorist attacks, detection of the onset of an epidemic, and failure detection in manufacturing systems and large machines. In all of these applications, the sensors monitoring the environment take observations that undergo a change in statistical distribution in response to the change in the environment. The goal is to detect this change point as quickly as possible, subject to false alarm constraints.

In the standard formulation of the change-point detection problem, there is a sequence of observations whose distribution changes at some unknown point in time and the goal is to detect this change as soon as possible, subject to false alarm constraints [1],[5]-[12]. In this paper, we are interested in the generalization of this problem that corresponds to the multi-sensor situation where the information available for decision-making is distributed (decentralized). The observations are taken at a set of \( N \) distributed sensors as shown in Figure 1. The statistical properties of the sensors’ observations change at the same time. The sensors may send either quantized versions of their observations or local decisions to a fusion center where a final decision is made based on all the sensor messages.

Here we deal with the Bayesian problem formulation in which the change point is treated as a random variable with a known prior distribution that is geometric. We provide an asymptotic analysis of two centralized detection procedures, the Shiryaev and Shiryaev-Roberts procedures, under quite general conditions that are not confined to the conventional iid assumption. We also present results on the asymptotic analysis of these procedures in the decentralized setting where sensors send quantized versions of the observations to the fusion center, which performs change detection. In addition, we present constructions of decentralized detection procedures that perform local change detection at the sensors, and show that these procedures have the same first-order asymptotic performance as the central-
2.1 Problem formulation

Asymptotic performance of certain processing capabilities at the sensors. This is applicable in scenarios where ample bandwidth is available for communication between the sensors and fusion center. At an unknown point in time \( \tau \), the conditional density of \( X_n \) given \( X^{n-1} \) is

\[
p_k(X_n | X^{n-1}) = p_0(X_n | X^{n-1}),
\]

while for \( n \geq k \)

\[
p_k(X_n | X^{n-1}) = \prod_{i=1}^{N} f_{i,n}^{(i)}(X_{i,n} | X_i^{n-1}).
\]

In mathematical terms, a centralized sequential change-point detection procedure is identified with a stopping time \( \tau \) for an observed sequence \( \{X_n\}_{n \geq 1} \), i.e. \( \tau \) is an extended integer-valued random variable, such that the event \( \{\tau \leq n\} \) belongs to the sigma-algebra generated by the first \( n \) observations from all the sensors. A false alarm is raised whenever the detection is declared before the change occurs, i.e. when \( \tau < \lambda \). A good detection procedure should guarantee a stochastically small detection delay \( \tau - \lambda \) provided that there is no false alarm (i.e. \( \tau \geq \lambda \)), while the rate of false positives should be low.

In a Bayesian setting, the change point \( \lambda \) is assumed to be random with prior probability distribution \( \pi_k = P(\lambda = k), \)

For \( k = 1, 2, \ldots \). The goal is to detect the change as soon as possible after it occurs, subject to constraints on the false alarm probability.

In what follows, \( P^\pi \) stands for the average probability measure, which is defined as

\[
P^\pi(\Omega) = \sum_{k=1}^{\infty} P_k(\Omega) \pi_k,
\]

and \( E^\pi \) denotes the expectation with respect to \( P^\pi \).

In the Bayesian setting, a reasonable measure of the detection lag is the average detection delay (ADD)

\[
ADD(\tau) = E^\pi(\tau - \lambda | \tau \geq \lambda)
\]

\[
= \sum_{k=1}^{\infty} \pi_k E_k(\tau - k | \tau \geq k),
\]

while the false alarm rate can be measured by the probability of false alarm

\[
PFA(\tau) = P^\pi(\tau < \lambda) = \sum_{k=1}^{\infty} \pi_k P_k(\tau < k).
\]

An optimal Bayesian detection procedure is a procedure for which ADD is minimized while PFA(\( \tau \)) is set at a given level \( \alpha, 0 < \alpha < 1 \). Specifically, define the class of change-point detection procedures \( \Delta(\alpha) = \{ \tau : PFA(\tau) \leq \alpha \} \).
for which the false alarm probability does not exceed the predefined number $\alpha$. The optimal change-point detection procedure is described by the stopping time

$$\nu = \arg \inf_{\tau \in \Delta(\alpha)} \text{ADD}(\tau).$$

Let $p_n = P(\lambda \leq n | X^n)$ be the posterior probability that the change occurred before time $n$. For the iid case, where $f^{(i)}_{0,0}(X_{i,0} | X_{i,j-1}) = f^{(i)}_{0,0}(X_{i,j})$ and $f^{(i)}_{1,0}(X_{i,1} | X_{i,j-1}) = f^{(i)}_{1,0}(X_{i,j})$, it follows from works of Shiryaev [8, 9] that if the distribution of the change point is geometric, then the optimal centralized detection procedure is the one that raises an alarm at the first time such that the posterior probability $p_n$ exceeds a threshold $A$,

$$\nu(A) = \inf \{ n \geq 1 : p_n \geq A \},$$

(3)

where the threshold $A = A_\alpha$ should be chosen in such a way that $\text{PFA}(\nu(A)) = \alpha$. However, except for the case of detecting the change in the drift of the Wiener process observed in continuous time it is difficult to find a threshold that provides an exact match to the given PFA. Also, until recently there were no results related to the ADD evaluation of this optimal procedure, again except for the continuous-time Wiener process.

While the exact match of the false alarm probability is related to the estimation of the overshoot in the stopping rule (3), and for this reason is problematic, a simple upper bound, which ignores overshoot, can be obtained [12]. Indeed, since $P^n \{ \nu(A) < \lambda \} = E^n \{ 1 - p_{\nu(A)}(\lambda) \}$ and $1 - p_{\nu(A)} \leq 1 - A$ on $\{ \nu(A) < \infty \}$, it follows that the PFA defined in (2) obeys the inequality

$$\text{PFA}(\nu(A)) \leq 1 - A.$$  

(4)

Thus, setting $A = A_\alpha = 1 - \alpha$ guarantees the inequality $\text{PFA}(\nu(A_\alpha)) \leq \alpha$. Note that inequality (4) holds true for arbitrary (proper), not necessarily geometric, prior distributions and for arbitrary non-iid models.

In the rest of the paper, we assume that the prior distribution of the change point is geometric with the parameter $\rho$, $0 < \rho < 1$, i.e.

$$\pi_k = P(\lambda = k) = \rho(1 - \rho)^{k-1} \text{ for } k = 1, 2, \ldots.$$  

For $k \leq n$, introduce the following two statistics

$$\Lambda_k := \frac{dP_k(X^n)}{dP_\infty(X^n)} = \prod_{i=k}^{n} \frac{f^{(i)}_{1,0}(X_{i,1} | X_{i,t-1})}{f^{(i)}_{0,0}(X_{i,0} | X_{i,t-1})}$$

(5)

and

$$R_{p,n} = \sum_{k=1}^{n} (1 - \rho)^{k-1-n} \Lambda_k.$$  

(6)

Taking into account that $R_{p,n} = p_n/(1 - p_n)\rho$, the Shiryaev stopping rule given in (3) can be written in the following form

$$\nu_B = \inf \{ n \geq 1 : R_{p,n} \geq B \}, \quad B = \frac{A}{(1 - A)\rho}.$$  

(7)

Consequently,

$$B_\alpha = (1 - \alpha)/(\alpha \rho) \implies \nu_{B_\alpha} \in \Delta(\alpha).$$

(8)

It is worth mentioning that while the Shiryaev procedure (7) is optimal in the iid case, it may not be optimal in the non-iid scenario even if we can set the threshold to meet the PFA constraint $\alpha$. Our recent study, however, shows that it is asymptotically optimal when $\alpha$ approaches zero under fairly general conditions [12].

In addition to the Bayesian Shiryaev procedure, we will also be interested in a related procedure, namely the Shiryaev-Roberts detection procedure (Shiryaev 1961, Roberts 1965). The Shiryaev-Roberts (SR) procedure is defined by the stopping time

$$\hat{\tau}_B = \inf \{ n : R_n \geq B \},$$

(9)

where the statistic $R_n$ is given by

$$R_n = \sum_{k=1}^{n} \Lambda_{k}.$$  

(10)

It is known that in the case where the observations are iid and the change point is modeled as deterministic but unknown, the SR procedure is asymptotically optimal with respect to the minimax expected detection lag, subject to a constraint on the mean time to false alarm [1, 7, 10]. Below we show that this change-point detection procedure loses the optimality property under the Bayesian criterion (in class $\Delta(\alpha)$).

In addition to the Bayesian ADD defined in (1), we will also analyze the behavior of the conditional ADD (CADD) for the fixed change point $\lambda = k$, which is defined by

$$\text{CADD}_k(\tau) = E_k(\tau - k | \tau \geq k) \quad k = 1, 2, \ldots.$$  

\footnote{The statistic $R_n = \lim_{\rho \to 0} R_{p,n}$.}

### 2.2 Asymptotic optimality of the Shiryaev procedure

For $i = 1, \ldots, N$, define the statistics

$$Z_{i,n}^k = \sum_{\ell=k}^{n} \log \frac{f^{(i)}_{1,0}(X_{i,1} | X_{i,t-1})}{f^{(i)}_{0,0}(X_{i,0} | X_{i,t-1})},$$

which can be interpreted as the log-likelihood ratios (LLR) of the hypotheses $H_k : \lambda = k$ (the change occurs at $\lambda = k$) and $H_\infty : \lambda = \infty$ (the change does not occur) in the

758
Theorem 1. Let $B = B_\alpha = (1-\alpha)/(\alpha p)$. Assume that conditions (11) and (12) hold for some positive $q_i$. Then, as $\alpha \to 0$,

$$\inf_{\tau \in \Delta^{(\alpha)}} \text{CADD}_k(\tau) \sim \text{CADD}_k(\nu_{B_\alpha}) \sim \frac{1}{q_{\text{tot}}} + \log(1-\rho),$$

for all $k \geq 1$;

$$\inf_{\tau \in \Delta^{(\alpha)}} \text{ADD}(\tau) \sim \text{ADD}(\nu_{B_\alpha}) \sim \frac{1}{q_{\text{tot}}} + \log(1-\rho),$$

(13) (14)

We stress that the detection procedure $\nu_B$ with the threshold $B = B_\alpha = (1-\alpha)/(\alpha p)$ is asymptotically optimal not only relative to the ADD, but also uniformly asymptotically optimal with respect to the conditional ADD for all values of $\lambda = k, k = 1, 2, \ldots$.

Next, consider the iid case where $f^{(i)}_{0,n}(X_{i,n}|X_i^{n-1}) = f^{(i)}_0(X_{i,n})$ and $f^{(i)}_{1,n}(X_{i,n}|X_i^{n-1}) = f^{(i)}_1(X_{i,n})$. The LLRs get modified to

$$Z_{i,n}^k = \sum_{t=k}^n \log f^{(i)}_0(X_{i,t}), \quad Z_{i,n}^\lambda = \sum_{t=k}^n \log f^{(i)}_1(X_{i,t}),$$

and the decision statistic $R_{\rho,n}$ satisfies the recursion

$$R_{\rho,n} = \frac{1}{1-\rho}(1 + R_{\rho,n-1}) \prod_{i=1}^n e^{Z_{i,n}}, R_{\rho,0} = 0, \quad (15)$$

which may be deployed for practical implementation and simulations.

As it was mentioned in Section 2.1, in the iid case the Shiryaev procedure (7) is optimal when the threshold $B$ can be chosen in such a way that $\text{PFA}(\nu_B) = \alpha$. Since it is difficult to meet this exact requirement, we will study the properties of the detection procedure $\nu_B$ with $B = (1-\alpha)/(\alpha p)$, which guarantees the inequality $\text{PFA}(\nu_B) \leq \alpha$.

Let

$$I_i = E_1 Z_{i,1}^1 = \int \log \left( \frac{f^{(i)}_1(x)}{f^{(i)}_0(x)} \right) f^{(i)}_1(x) dx$$

be the Kullback-Leibler (K-L) information number between the densities $f^{(i)}_1(x)$ and $f^{(i)}_0(x)$. In the iid case, the K-L numbers $I_i$ play the role of the numbers $q_i$ that appeared in Theorem 1.

In the iid case, the second moment condition $E[I^2|Z_{i,1}^1]^2 < \infty$ is both necessary and sufficient for the complete (average complete) convergence of $Z_{i,n}^1/(n-\lambda + 1)$ to $I_i$, which follows from the Baum-Katz rates of convergence in the law of large numbers [2].

In the iid case, the second moment condition $E[I^2|Z_{i,1}^1]^2 < \infty$ is both necessary and sufficient for the complete (average complete) convergence of $Z_{i,n}^1/n$ to $I_i$, which follows from the Baum-Katz rates of convergence in the law of large numbers [2].

Theorem 5 in [12], the Shiryaev procedure minimizes not only ADD but also all positive moments of the detection delay whenever K-L numbers $I_i$ are positives and finite.

An asymptotically accurate approximation for the PFA and higher order approximations for the ADD that take into account overshoots can be obtained in just the same way as in [12], Theorem 6.
2.3 Asymptotic performance of the Shiryaev-Roberts procedure

In this section, we analyze the behavior of the SR change-point detection procedure defined above in (9) and (10).

In order to obtain an upper bound for PFA of the SR procedure, we note that the statistic $R_n$ is a $P_\infty$-submartingale with mean $E_\infty R_n = n$. Using the Doob’s submartingale inequality, we get

$$P_\infty \{ \hat{\tau}_B < n \} = P_\infty \left\{ \max_{1 \leq k < n} R_k \geq B \right\} \leq n/B,$$

which yields

$$\text{PFA}(\hat{\tau}_B) = \sum_{n=1}^{\infty} \pi_n P_\infty \{ \hat{\tau}_B < n \} \leq 1/(\rho B).$$

Thus, choosing $B = 1/(\alpha \rho)$ guarantees $\hat{\tau}_{B_0} \in \Delta(\alpha)$.

The following theorem determines the asymptotic performance of the SR procedure. Its proof will be given elsewhere. See, however, Theorem 5 in [12] for the iid case.

**Theorem 2.** Let $B_0 = A_0 = 1/(\alpha \rho)$. Assume that conditions (11) and (12) hold for some positive $q_i$. Then, as $\alpha \to 0$,

$$\text{CADD}_k(\hat{\tau}_{B_0}) \sim \frac{|\log \alpha|}{q_{\text{tot}}} \quad \text{for all } k \geq 1; \quad (17)$$

$$\text{ADD}(\hat{\tau}_{B_0}) \sim \frac{|\log \alpha|}{q_{\text{tot}}}. \quad (18)$$

Comparing Theorem 1 and Theorem 2 shows that SR detection procedure is not asymptotically optimal in the class $\Delta_1(\alpha)$ (under the Bayesian constraint). In particular, the relative asymptotic efficiency

$$\lim_{\alpha \to 0} \frac{\text{ADD}(\hat{\tau}_{B_0})}{\inf_{\tau \in \Delta(\alpha)} \text{ADD}(\tau)} = 1 + \frac{|\log(1 - \rho)|}{q_{\text{tot}}}. \quad (19)$$

Clearly the SR detection procedure is close to optimum when $q_{\text{tot}} \gg |\log(1 - \rho)|$. Interestingly, in practice, the SR detection procedure performs close to the optimal procedure even when this condition does not hold, as long as $\alpha$ is not too small (see Section 5).

3 Decentralized detection with quantizers at the sensors

In this section, we discuss the asymptotic performance and asymptotic optimality results of the decentralized version of the change detection problem described in the work by Veeravalli [14].

It is supposed that based on the information available at the sensor $S_i$ at time $n$, a message $U_{i,n}$, belonging to a finite alphabet of size $V_i$, is formed and sent to the fusion center (see Figure 1). We will use the vector notation: $X_n = (X_{1,n}, \ldots, X_{N,n})$ and $U_n = (U_{1,n}, \ldots, U_{N,n})$.

Based on the sequence of sensor messages, a decision about the change is made at the fusion center. The fusion center picks a time $\tau$, which is a stopping time on $\{U_n\}_{n \geq 1}$, at which it is declared that a change has occurred. For the sake of simplicity, we restrict our attention only to the iid case. A generalization to the non iid case can be obtained by using the same kind of argument as in Subsection 2.2.

Various information structures are possible for the decentralized configuration depending on how feedback and local information is used at the sensors [14]. Consider the simplest information structure where the message $U_{i,n}$ formed by sensor $S_i$ at time $n$ is a function of only its current observation $X_{i,n}$, i.e., $U_{i,n} = \psi_{i,n}(X_{i,n})$.

When the observations at each sensor are iid, for stationary sensor quantizers the K-L distances between the quantizers at the sensors, an application of Theorem 1 gives us that $g_j^{(i)}$ are positive and finite, then for fixed stationary sensor quantizers, an application of Theorem 1 gives us that $\nu_B = \inf \{ n \geq 1 : R_{p,n} \geq B \}$,

where $B$ is a positive threshold which is selected so that PFA($\nu_B$) $\leq \alpha$.

If $I_i = I_i(g_1^{(i)}, g_0^{(i)})$, the K-L distances between the $g_1^{(i)}$ and $g_0^{(i)}$, are positive and finite, then for fixed stationary sensor quantizers, an application of Theorem 1 gives us that...
the detection procedure $\nu_B$ given in (20), with $B = B_\alpha = (1 - \alpha)/(\alpha \rho)$, is asymptotically optimal as $\alpha \to 0$ among all procedures with PFA no greater than $\alpha$. To be specific, let $\psi = \{\psi_1, \ldots, \psi_N\}$ be a stationary quantizer. Then, as $\alpha \to 0$,

$$
\inf_{\tau \in \Delta(\alpha)} \text{ADD}(\psi, \tau) \sim \text{ADD}(\nu_B, \psi)
$$

$$
\sim \frac{|\log \alpha|}{|\log(1 - \rho)| + \sum_{i=1}^N I_i(g^{(i)}_1, g^{(i)}_0)},
$$

where $\text{ADD}(\psi, \tau) = E(\tau - \lambda | \tau \geq \lambda)$ is the ADD for the policy $(\psi, \tau)$.

This result immediately reveals how to optimize the sensor quantizers: It is asymptotically optimum (as $\alpha \to 0$) for sensor $S_i$ to use the stationary quantizer $\psi_{i, \text{opt}} = \arg \max_{i, \tau} I_i(g^{(i)}_1, g^{(i)}_0)$ that maximizes the K-L information distance.

Based on the results of Tsitsiklis [13], it is easy to show that the optimal stationary quantizer $\psi_{i, \text{opt}}$ is a monotone likelihood ratio quantizer (MLRQ), i.e. there exist thresholds $\beta_{i,1}, \beta_{i,2}, \ldots, \beta_{i,v_i-1}$ satisfying $0 = \beta_{i,0} \leq \beta_{i,1} \leq \beta_{i,2} \leq \cdots \leq \beta_{i,v_i-1} \leq \infty = \beta_{i,v_i}$ such that, for $j = 1, \ldots, V_i$,

$$
\psi_{i, \text{opt}}(X) = j \text{ only if } \beta_{i,j-1} < \frac{f^{(i)}_1(X)}{f^{(i)}_0(X)} \leq \beta_{i,j}.
$$

Thus, the asymptotically optimal policy $\phi_{\text{opt}}$ for a decentralized change detection problem in the class of stationary (in time) quantizers consists of a set of MLRQ’s at the sensors followed by the Shiryaev’s procedure based on $\{U_{n}\}_{n \geq 1}$ at the fusion center (as described in (20)).

For each $i$, let the probability induced on $U_i,n$ by the optimal MLRQ $\psi_{i, \text{opt}}$ be given by $g^{(i)}_{1, \text{opt}}$ and $g^{(i)}_{0, \text{opt}}$. Then the effective K-L information distance between the ‘change’ and ‘no change’ hypotheses at the fusion center is given by

$$
I_{i, \text{opt}} = \sum_{i=1}^N I_i(g^{(i)}_{1, \text{opt}}, g^{(i)}_{0, \text{opt}}).
$$

Finally, we denote by $\nu_{\text{opt}}$ the stopping rule at the fusion center for the case where the sensor quantizers are chosen to be $\psi_{i, \text{opt}}$. and by $\Phi_{\text{st}}(\alpha)$ the class of policies $\phi$ with all stationary quantizers and stopping rules at the fusion center such that $\tau \in \Delta(\alpha)$.

The asymptotic performance of the asymptotically optimum solution to the decentralized change detection problem described above is given in the following theorem, which follows directly from Theorem 1 and the argument given above.

**Theorem 3.** Suppose that

$$
0 < I_i(g^{(i)}_{1, \text{opt}}, g^{(i)}_{0, \text{opt}}) < \infty \text{ for } i = 1, \ldots, N.
$$

Then $B_\alpha = (1 - \alpha)/(\alpha \rho)$ implies that $PFA(\nu_{\text{opt}}) \leq \alpha$ and

$$
\inf_{\phi \in \Phi_{\text{st}}(\alpha)} \text{ADD}(\phi) \sim \text{ADD}(\phi_{\text{opt}})
$$

$$
\sim \frac{|\log \alpha|}{I_{i, \text{opt}} + |\log(1 - \rho)|} \text{ as } \alpha \to 0.
$$

It follows from Theorems 1 and 3 that the relative asymptotic efficiency of optimal centralized and decentralized detection procedures is equal to

$$
E_{\text{opt}} = \lim_{\alpha \to 0} \frac{\text{ADD}(\nu_{B_\alpha})}{\text{ADD}_{\text{dc}}(\nu_{B_\alpha})} = \frac{I_{i, \text{opt}} + |\log(1 - \rho)|}{I_{i, \text{opt}} + |\log(1 - \rho)|}.
$$

Since $I_{i, \text{opt}}$ is always larger than $I_{i, \text{opt}}$, the value of $E_{\text{opt}}$ is always smaller than 1, and this is always the case if $I_{i, \text{opt}} \gg |\log(1 - \rho)|$.

In the next section, we give constructions of two decentralized detection procedures (with sensors performing local change detection) for which the first-order asymptotic performance is the same as that for the centralized SR procedure.

### 4 Decentralized detection with local change detection at the sensors

The results of the previous section show that fusion of data in decentralized multi-sensor systems with quantizers always leads to certain losses of information which results in the performance degradation of the optimal decentralized policy. We now explore the scenario where the sensors perform local change detection to see if this loss in performance can be eliminated asymptotically.

Let $R_n^{(i)}$ be the SR statistic in the sensor $S_i$ based on the original, non-quantized data $X_n = (X_{i,1}, \ldots, X_{i,n})$ (see (10)) and introduce the local stopping times in the sensors,

$$
\tau_i = \inf \left\{ n : R_n^{(i)} \geq B_i \right\}, \quad i = 1, \ldots, N.
$$

We propose two rules to fuse the sensor decisions at the fusion center.

For the first fusion rule, at time $n$, local binary decisions $d_{i,n} = 0$ or 1 are transmitted to the fusion center, where $d_{i,n} = 1$ if $R_n^{(i)} \geq B_i$, and 0 otherwise. In this case, the stopping time at the fusion center is given by:

$$
\tau_{\text{all}} = \text{first } n \geq 1 \text{ such that } d_{i,n} = 1 \text{ for all } i = 1, \ldots, N.
$$

$$
(22)
$$
Note that stopping time can be rewritten as:

$$\tau_{all} = \inf \left\{ n : \min_{1 \leq i \leq N} \left( R_n^{(i)} / B_i \right) \geq 1 \right\}. \quad (23)$$

For the second fusion rule, we stop monitoring the \(i\)-th sensor once the exceedance has occurred and transmit the decision \(d_i = 1\) at time \(\tau_i\) to the fusion center. At the fusion center the decision in favor of the change hypothesis is made once all the sensors “vote” for this hypothesis. This fusion procedure is equivalent to the stopping time

$$\tau_{max} = \max_{1 \leq i \leq N} \tau_i. \quad (24)$$

Note that \(\tau_{all}\) is greater than \(\tau_{max}\) (almost surely).

To bound the PFA of these procedures, we use the following steps. Since, under \(P_{\infty}\), the local stopping times \(\tau_1, \ldots, \tau_N\) are independent and, by (16)

$$P_{\infty}(\tau_i < n) \leq n / B_i, \quad n \geq 1,$$

setting \(B_i = h^{a_i}\) with \(a_i = q_i / q_{tot}\) and \(h > 0\), we obtain

$$PFA(\tau_{max}) = \sum_{n=1}^{\infty} \pi_n \prod_{i=1}^{N} P_{\infty}(\tau_i < n) \leq \sum_{i=1}^{N} h^{-a_i} \sum_{n=1}^{\infty} n^{N} \pi_n,$$

Noting that \(\sum_{i=1}^{N} q_i = 1\) and \(\sum_{n=1}^{\infty} n^{N} \pi_k = \beta_N\) yields

$$PFA(\tau_{max}) \leq \beta_N / h,$$

where \(\beta_N\) is the \(N\)-th moment of the geometric random variable with parameter \(p\). Therefore, \(h_{\alpha} = \beta_N / \alpha\) guarantees the inequality \(PFA(\tau_{max}) \leq \alpha\).

Clearly the same bound holds for \(\tau_{all}\) as well, since \(\tau_{all}\) is greater than \(\tau_{max}\) almost surely.

Furthermore, it can be shown that in the conditions of Theorems 1 and 2

$$ADD(\tau_{max}) \sim ADD(\tau_{all}) \sim \max_{1 \leq i \leq N} \frac{\log B_i}{q_i},$$

as \(\min_i B_i \to \infty\). Since

$$\log h_{\alpha} = | \log \alpha | + \log \beta_N = | \log \alpha | + O(1),$$

it follows that

$$ADD(\tau_{max}) \sim ADD(\tau_{all}) \sim \frac{| \log \alpha |}{q_{tot}} \quad \text{as} \alpha \to 0.$$

The following theorem formalizes the asymptotic performance of the proposed decentralized detection procedures.

**Theorem 4.** Let \(B_i = (\beta_N / \alpha)^{n_i / q_{tot}}, i = 1, \ldots, N\) and assume that conditions (11) and (12) hold for some positive \(q_i\). Then, as \(\alpha \to 0\),

$$CADD_k(\tau_{max}) \sim CADD_k(\tau_{all}) \sim \frac{| \log \alpha |}{q_{tot}} \quad \forall k,$$

$$ADD(\tau_{max}) \sim ADD(\tau_{all}) \sim \frac{| \log \alpha |}{q_{tot}}. \quad (25)$$

Comparing (17) and (18) in Theorem 2 with (25), we can see that the first-order asymptotic operating characteristics of the proposed decentralized procedures are the same as those of the SR centralized procedure in the sense that

$$\lim_{\alpha \to 0} \frac{ADD(\tau_{max})}{ADD_c(\hat{\tau})} = \lim_{\alpha \to 0} \frac{ADD(\tau_{all})}{ADD_c(\hat{\tau})} = 1.$$

However, these decentralized procedures have somewhat worse second-order performance, since for the SR centralized procedure the second term of expansion of ADD is a constant, while for \(\tau_{max}\) and \(\tau_{all}\) it grows as a square root of the threshold. Thus,

$$ADD(\tau_{max}) - ADD_c(\hat{\tau}) = O(| \log \alpha |^{1/2}) \to \infty$$

as \(\alpha \to 0\).

We remark that we were unable to obtain similar results if the optimal Bayesian (Shiryaev) procedure is used in place of the SR procedure at the sensors. We also note that while the procedures \(\tau_{max}\) and \(\tau_{all}\) are asymptotically almost globally optimal, their performance for moderate values of \(\alpha\) may be far from optimum, and may even be inferior to the procedure that uses binary quantizers at the sensors (see Figure 2). Finally, since \(\tau_{all}\) uses more information than \(\tau_{max}\), we expect it to perform better.

### 5 An example

Consider the problem of detecting a non-fluctuating target using \(N\) geographically separated sensors. The observations are corrupted by additive white Gaussian noise that is independent from sensor to sensor. The sensors preprocess the observations using a matched filter, matched to the signal corresponding to the target. The output of the matched filter at sensor \(S_i\) at time \(n\) (when the time of appearance of the target is \(\lambda\)) is given by:

$$X_{i,n} = \xi_{i,n},$$

and

$$X_{i,n} = \mu_i + \xi_{i,n},$$

where \(\{\xi_{i,n}, n = 1, 2, \ldots\}\) is a sequence of i.i.d. zero-mean Gaussian random variables with variance \(\sigma_i^2\). Therefore, the likelihood ratio at sensor \(S_i\) is given by

$$Y_i(x) = \frac{f_i^{(1)}(x)}{f_i^{(0)}(x)} = \exp \left\{ \frac{\mu_i (x - \mu_i / 2)}{\sigma_i^2} \right\}.\quad (26)$$

Note that the likelihood ratio is monotonically increasing, and we can hence characterize the optimal stationary sensor quantizers in terms of thresholds on the observations. For binary decisions at the sensors, the quantizers are characterized by a single threshold, i.e. \(U_{i,n} = 1\) if \(X_{i,n} \geq \beta_i\) and 0 otherwise. The distributions induced on \(U_{i,n}\) by this quantizer are given by:

$$g_{j_{i}}^{(i)}(0) = 1 - g_{j_{i}}^{(i)}(1) = \Phi \left( \frac{\beta_i - j_{i} \mu_i}{\sigma_i} \right) = q_{j_{i}}^{(i)}, \quad j = 0, 1.$$

\(^{2}\)A proof of the latter fact will be given elsewhere.
The optimal value of $\beta_i$, i.e., the one that maximizes $L_i(g_1^{(i)}, g_0^{(i)})$, is easily found based on this formula. Then we can compute the decision statistics $R_{p,n}$ and $R_n$ for the fusion center as well as the statistic $R_{i,n}^{(i)}$ for the sensor $S_i$.

The operating characteristics in an example with three sensors having identically distributed observations are illustrated in Figure 2. The parameter values are $\rho = 0.1$, $\mu_t = 0.4$ and $\sigma_t^2 = 1$. The K-L distance for the sensor observations is 0.08. The threshold that maximizes the K-L distance at the output of the sensor is $\beta = 0.32$, and the corresponding maximum K-L distance is 0.0509. Estimates of PFA and ADD are obtained using MC methods with the number of trials being $1000/\alpha$. As we expect, for the optimal (Shiryaev) centralized policies, the plot of ADD versus $-\log(PFA)$ is roughly a straight line with slope that is approximately equal to $1/[3I(f_1, f_0) + \log(1 - \rho)] \approx 2.89$. Interestingly, the SR centralized policy has very similar performance even though the asymptotic slope in this case is $1/[3I(f_1, f_0)] \approx 4.17$. This justifies the use of the SR policy at the sensors in constructing $\tau_{\text{max}}$ and $\tau_{\text{all}}$. The decentralized policy with sensors quantizing their observations to one bit has a tradeoff curve with slope that is roughly equal to $1/[\ln(c) + \log(1 - \rho)] \approx 3.87$ as expected from Theorem 1. The decentralized policy of course suffers a performance degradation relative to the centralized policy. However, the bandwidth requirements for communication with the fusion center are considerably smaller in decentralized setting, especially with binary quantizers.

Figure 2 also shows the tradeoff curves for the procedures $\tau_{\text{max}}$ and $\tau_{\text{all}}$, where the sensors perform local change detection. As expected from the analysis in Section 4, $\tau_{\text{max}}$ performs worse than $\tau_{\text{all}}$. Further, it is interesting to see that both of these procedures have performances that are far from that of the centralized SR procedure. Thus, the asymptotic results of Theorem 4 appear to hold only for PFA much smaller than $10^{-4}$ (the smallest value of PFA considered in our simulations\(^3\)). In particular, $\tau_{\text{max}}$ performs even worse than the decentralized policy with binary quantizers, while $\tau_{\text{all}}$ is slightly better than the decentralized policy for sufficiently small PFA. These results clearly point to the need for further research on designing procedures that perform local detection at the sensors.

References


\(^3\)Note that Figure 2 shows the natural logarithm of PFA.