

Chapter 17

Change-Point Detection in Multichannel and Distributed Systems

ALEXANDER G. TARTAKOVSKY

University of Southern California, Los Angeles, USA

VENUGOPAL V. VEERAVALLI

University of Illinois at Urbana-Champaign, Urbana, USA

17.1 INTRODUCTION

The goal of this paper is to show that recent advances in change-point detection theory, as described in Basseville and Nikiforov (1993), Dragalin (1995,1996), Lai (1995,1998), Pollak (1985,1987), Tartakovsky (2003), and others, can be successfully applied to (i) certain practical problems related to rapid detection of targets in multichannel and multisensor distributed systems, and (ii) the problem of building high-speed anomaly detection systems for early detection of intrusions in large-scale distributed computer networks. We show that the asymptotic theory that has been developed for change-point detection is useful in practical engineering problems too, and it allows for the development of efficient algorithms that are easily implemented. In addition to that, these algorithms have certain optimality properties.

In the standard formulation of a change-point detection problem, one assumes having a sequence of observations whose distribution changes at some unknown instant λ , $\lambda = 1, 2, \dots$. The goal is to detect this change as soon as possible, subject to false alarm constraints. We are interested in two generalizations of this problem.

The first generalization refers to *multichannel* systems, where there are N sequences of observations whose distributions follow a certain law up to some unknown instant λ , $\lambda = 1, 2, \dots$. After this instant, one of the populations (and only one) changes its statistical properties. We wish to detect the change as soon as possible after it occurs, subject to constraints on the rate of false alarms. We assume that it is not necessary to indicate which channel has changed. Only the fact that a change occurred is important.

The second generalization corresponds to a situation where the information available for decision-making is distributed (or decentralized). Here, the observations are made at a set of L distributed sensors. The sensors' observations could in general be multichannel, and at the change-point λ , one channel at *each* sensor could change distribution. The sensors send quantized versions of their observations to a *fusion center* where change detection is performed based on the messages from all sensors.

These generalizations are of considerable practical importance and they arise in a variety of applications such as biomedical signal processing, quality control engineering, finance, link failure detection in communication networks, intrusion detection in computer networks, and target detection in surveillance systems. In particular, a typical scenario involves the detection of a target which appears randomly at an unknown time in an N -channel system (infrared, radar, sonar, etc.). It is necessary to detect the target "as quickly as possible" while maintaining the false alarm rate at a given level. Another important application area is intrusion detection in distributed high-performance computer networks. Large scale attacks, such as external denial-of-service attacks or internal man-in-the-middle attacks, occur at unknown points in time and should be detected in the early stages by observing abrupt (usually small) changes in the network traffic compared to the "normal" (legitimate) mode. These application areas will be emphasized in Section 17.4.

17.2 MULTICHANNEL CHANGE DETECTION: THEORY

17.2.1 Problem Formulation

We assume that the observed stochastic process $\mathbf{X}_k = (X_{1,k}, \dots, X_{N,k})$ is an N -component process. The component $X_{i,k}$, $k = 1, 2, \dots$, corresponds to observations obtained from the i^{th} channel of an N -channel system, and all of the channels can be observed simultaneously. Let \mathbf{P}_∞ stand for the probability measure when the change does not occur ($\lambda = \infty$), and let $\mathbf{P}_{\lambda,i}$ be the probability measure when the change occurs at time λ in the i -th channel. Note that if $\lambda = \infty$, then $\mathbf{P}_{\infty,i} = \mathbf{P}_\infty$ for all i . Further, let \mathbf{E}_∞ and $\mathbf{E}_{\lambda,i}$ denote the corresponding expectations. A sequential change-point detection procedure is described by a random time τ depending on the observations, which is a *stopping time* (with respect to the family of sigma-algebras $\mathcal{F}_k = \sigma(\mathbf{X}_1, \dots, \mathbf{X}_k)$, $k \geq 0$) at which it is declared that a change has occurred. Typically, this stopping time (time of alarm) is defined as a first time $k \geq 1$ when some statistic exceeds a threshold that controls the rate of false alarms.

Designing the quickest change detection procedures usually involves optimizing the tradeoff between two kinds of performance measures, one being a measure of detection delay, and the other being a measure of the frequency of false alarms. There are two standard mathematical formulations for the optimum tradeoff problem. The first of these is a minimax formulation, due to Lorden (1971) and Pollak (1985), in which the goal is to minimize the worst-case delay, $\text{ES}_i(\tau) = \sup_\lambda \text{ess sup } \mathbf{E}_{\lambda,i}[(\tau - \lambda + 1)^+ | \mathbf{X}_1, \dots, \mathbf{X}_{\lambda-1}]$ or $D_i(\tau) = \sup_\lambda \mathbf{E}_{\lambda,i}(\tau - \lambda | \tau \geq \lambda)$, subject to a lower bound on the mean time between false alarms ("ess sup" of a set being the supremum of it except possibly a subset of measure zero). The second is a Bayesian formulation, proposed by Shiryaev (1963,1978), 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 the false alarm probability.

In what follows, we consider a minimax approach with Pollak's measure $D_i(\tau)$, or with a more general measure

$$D_i^m(\tau) = \sup_{1 \leq \lambda < \infty} \mathbf{E}_{\lambda,i}\{(\tau - \lambda)^m | \tau \geq \lambda\} \quad \text{for } m > 0, \quad (17.2.1)$$

which is nothing but the m^{th} moment of the detection delay, in the worst case, assuming that the change occurs in the i^{th} channel. Indeed,

if we show that the values of $D_i^m(\tau)$ are relatively small for all $1 \leq i \leq N$ and $m > 0$, then we will have shown that the entire distribution $\mathbf{P}_{\lambda,i}(\tau - \lambda | \tau \geq \lambda)$ is concentrated close to the change-point.

The constraint imposed on the false alarms is that the mean time to false alarm, $\mathbf{E}_\infty(\tau)$, should exceed a predefined number $T > 0$. In other words, we are interested in the class of detection procedures $\Delta(T) = \{\tau : \mathbf{E}_\infty(\tau) \geq T\}$. It is very difficult, if not impossible, to find an optimal procedure that minimizes the “risk functions” (17.2.1) in the class $\Delta(T)$ for an arbitrary value of T . For this reason, we will consider the asymptotic setting as $T \rightarrow \infty$, that is, we are interested in the following problem: Compute

$$\inf_{\tau \in \Delta(T)} D_i^m(\tau) \text{ as } T \rightarrow \infty \text{ for all } i = 1, \dots, N.$$

When we evaluate the performance of a change detection procedure in particular problems and examples, we will be interested in the average detection delay (ADD) and the false alarm rate (FAR), which are defined by:

$$\text{ADD}(\tau) = \frac{1}{N} \sum_{i=1}^N D_i(\tau), \quad \text{FAR}(\tau) = \frac{1}{\mathbf{E}_\infty(\tau)}. \quad (17.2.2)$$

17.2.2 The Detection Procedure and False Alarm Rate

Write $\mathbf{X}_i^n = (X_{i,1}, \dots, X_{i,n})$ and $\mathbf{X}^n = (\mathbf{X}_1^n, \dots, \mathbf{X}_N^n)$ for the concatenation of the first n observations from the i^{th} channel, and from all N channels, respectively. We suppose that the data in the channels, that is the vectors $\mathbf{X}_1^n, \dots, \mathbf{X}_N^n$, are mutually independent. However, in general, we do not assume that the data in a particular channel, say $X_{i,1}, X_{i,2}, \dots$, are i.i.d. before or after the change.

Let \mathbf{P}_∞ stand for the probability measure under which the conditional density of \mathbf{X}_k given $\mathbf{X}^{k-1} = \mathbf{x}^{k-1}$ is

$$f_{0,k}(\mathbf{x}_k | \mathbf{x}^{k-1}) = \prod_{\ell=1}^N p_{0,k}(x_{\ell,k} | \mathbf{x}_\ell^{k-1})$$

for every $k \geq 1$. For any $1 \leq \lambda < \infty$, we use $\mathbf{P}_{\lambda,i}$ to denote the probability measure under which the conditional density of \mathbf{X}_k given

$\mathbf{X}^{k-1} = \mathbf{x}^{k-1}$ is

$$f_{i,k,\lambda}(\mathbf{x}_k | \mathbf{x}^{k-1}) = \begin{cases} \prod_{\ell=1}^N p_{0,k}(x_{\ell,k} | \mathbf{x}_\ell^{k-1}) & \text{if } \lambda > k \\ p_{i,k}(x_{i,k} | \mathbf{x}_i^{k-1}) \prod_{\ell=1, \ell \neq i}^N p_{0,k}(x_{\ell,k} | \mathbf{x}_\ell^{k-1}) & \text{if } \lambda \leq k. \end{cases}$$

In other words, if the change occurs in the i^{th} channel at the time $\lambda = n$, the conditional probability density function (p.d.f.) of the i^{th} component changes from $p_{0,n}(x | \mathbf{x}_i^{n-1})$ to $p_{i,n}(x | \mathbf{x}_i^{n-1})$.

Next, let $H_{\lambda,i}$ be the hypothesis that the change occurs in the i^{th} channel at time $\lambda \in \{1, 2, \dots\}$, and let H_∞ be the hypothesis that the change does not occur at all (that is, $\lambda = \infty$). Then the *log-likelihood ratio* (LLR) between the hypotheses $H_{\lambda,i}$ and H_∞ is

$$Z_i^\lambda(n) := \log \frac{d\mathbf{P}_{\lambda,i}}{d\mathbf{P}_\infty}(\mathbf{X}^n) = \sum_{k=\lambda}^n \log \frac{p_{i,k}(X_{i,k} | \mathbf{X}_i^{k-1})}{p_{0,k}(X_{i,k} | \mathbf{X}_i^{k-1})}, \quad \lambda \leq n. \quad (17.2.3)$$

For $i = 1, \dots, N$, define the statistics $R_i(n) = \sum_{\lambda=1}^n e^{Z_i^\lambda(n)}$, and then combine these statistics to form the mixture

$$R(n) = N^{-1}(R_1(n) + \dots + R_N(n)).$$

The detection procedure $\tau^* = \tau^*(A)$ is defined as

$$\tau^*(A) = \min \{n \geq 1 : R(n) \geq A\} \quad (\tau^* = \infty \text{ if no such } n \text{ exists}),$$

where A is a positive number (threshold) that is selected so that the FAR is no larger than $1/T$. In other words, the moment of alarm is the first time n such that the statistic $R(n)$ exceeds the threshold A .

Note that the statistic $R_i(n)$ is the Shiryaev-Roberts statistic for detecting a change in the i^{th} channel (see, for example, Pollak (1985,1987)). The detection procedure τ^* is therefore an extension of the Shiryaev-Roberts procedure adapted to detect changes in multichannel systems.

We begin with an examination of the detection procedure τ^* under the hypothesis H_∞ . The following lemma establishes a simple lower bound for the average run length to false alarm $\mathbf{E}_\infty(\tau^*(A))$ in a general case. In Subsection 17.2.3, this bound will be improved in the i.i.d. case.

Lemma 17.2.1 *For an arbitrary stochastic model and any $A > 0$*

$$\frac{1}{\text{FAR}} = \mathbf{E}_\infty \tau^*(A) \geq A. \quad (17.2.4)$$

Proof: Since $\mathbf{E}_\infty\{e^{Z_i^n(n)} \mid \mathbf{X}^{n-1}\} = 1$, it is easily verified that, for any i , the statistics $R_i(n) - n$, $n \geq 1$, are \mathbf{P}_∞ -martingales with mean zero. This implies that the statistic $\{R(n) - n\}_{n \geq 1}$ is also a \mathbf{P}_∞ -martingale with a zero mean. Therefore, inequality (17.2.4) follows from the optional stopping theorem, which yields $\mathbf{E}_\infty(R(\tau^*)) = \mathbf{E}_\infty(\tau^*)$, and the fact that $R(\tau^*) \geq A$. Note that the optional stopping theorem can be applied because

$$\lim_{n \rightarrow \infty} \int_{\{\tau^* > n\}} |R(n) - n| d\mathbf{P}_\infty = 0.$$

It follows from the fact that $0 \leq R(n) < A$ on the set $\tau^* > n$. ■

It immediately follows from (17.2.4) that, by setting $A = T$, we guarantee the inequality $\mathbf{E}_\infty(\tau^*) \geq T$, that is, $A = T$ implies $\tau^*(T) \in \Delta(T)$.

17.2.3 Asymptotic Performance for Low FAR

In this subsection, we study the behavior of the detection procedure τ^* for large values of A and T , that is, for small FAR. In order to obtain asymptotic expansions for moments of the detection delay, we assume that the normalized LLRs

$$\frac{1}{n} Z_i^\lambda(n + \lambda - 1) = \frac{1}{n} \sum_{k=\lambda}^{n+\lambda-1} \log \frac{p_{i,k}(X_{i,k} \mid X_{i,1}, \dots, X_{i,k-1})}{p_{0,k}(X_{i,k} \mid X_{i,1}, \dots, X_{i,k-1})}, \quad i = 1, \dots, N,$$

converge almost surely (a.s.) as $n \rightarrow \infty$ to positive finite numbers J_i under $\mathbf{P}_{\lambda,i}$. Furthermore, we impose the following conditions on the rate of convergence:

$$\sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_{\lambda,i} \{ |Z_i^\lambda(n + \lambda - 1) - nJ_i| > n\varepsilon \} < \infty \quad \text{for all } \varepsilon > 0, \quad (17.2.5)$$

where r is a positive real number. If (17.2.5) holds, we say that $n^{-1} Z_i^\lambda(n + \lambda - 1)$ converges r -quickly to J_i under $\mathbf{P}_{\lambda,i}$ as $n \rightarrow \infty$ (Lai (1976)). For $r = 1$, (17.2.5) is the so-called complete convergence condition introduced by Hsu and Robbins (1947) in connection with the rates of convergence in the strong law of large numbers.

The following theorem establishes the asymptotic performance of the detection procedure $\tau^*(A)$ for large values of A . The proof is given in the Appendix. Hereafter we use the standard notation $X_A \sim Y_A$,

which means that $X_A/Y_A \rightarrow 1$ as $A \rightarrow \infty$.

Theorem 17.2.1 *Let condition (17.2.5) hold for some $r \geq 1$. Then for all $m \leq r$, $1 \leq \lambda < \infty$ and $i = 1, \dots, N$,*

$$\mathbf{E}_{\lambda,i} \{(\tau^*(A) - \lambda)^m | \tau^*(A) \geq \lambda\} \sim \left(\frac{\log NA}{J_i}\right)^m \quad \text{as } A \rightarrow \infty. \quad (17.2.6)$$

In addition, under some uniform conditions where (17.2.5) is strengthened to \sup_λ , first-order asymptotic optimality can be established with respect to “risks” $D_i^m(\tau^*)$ defined in (17.2.1). More precisely, under certain conditions, if $A = T$, then as $T \rightarrow \infty$,

$$\inf_{\tau \in \Delta(T)} D_i^m(\tau) \sim D_i^m(\tau^*) \sim \left(\frac{\log NT}{J_i}\right)^m, \quad i = 1, \dots, N. \quad (17.2.7)$$

Note, however, that in a general non-i.i.d. case the supremum in (17.2.1) is attained for some unspecified point λ that might even depend on T and go to infinity when $T \rightarrow \infty$.

Now consider the i.i.d. case assuming that, if the change occurs in the i^{th} population, then the observations $X_{i,1}, X_{i,2}, \dots$ are i.i.d. before the change with the p.d.f. $p_0(x)$ and after the change with the p.d.f. $p_i(x)$ (with respect to a sigma-finite measure $\mu(x)$). For the sake of brevity, in the rest of this subsection, we omit the index λ when $\lambda = 1$. For instance, we simply write $Z_i(n)$, \mathbf{E}_i , and \mathbf{P}_i instead of $Z_i^1(n)$, $\mathbf{E}_{1,i}$, and $\mathbf{P}_{1,i}$, respectively.

The first important observation is that $D_i^m(\tau^*) = \mathbf{E}_i(\tau^* - 1)^m$. Therefore, in further calculations related to the “risk” $D_i^m(\tau^*)$, we can concentrate on the evaluation of $\mathbf{E}_i(\tau^* - 1)^m$, assuming that the change occurs at the point $\lambda = 1$.

Define

$$\begin{aligned} \nu_i(a) &= \min \{n \geq 1 : Z_i(n) \geq a\} \quad (\nu_i(a) = \infty \text{ if no such } n \text{ exists}), \\ \gamma_i &= \lim_{a \rightarrow \infty} \mathbf{E}_i \exp \{-[Z_i(\nu_i(a)) - a]\}, \quad \bar{\gamma}_N = \frac{1}{N}(\gamma_1 + \dots + \gamma_N), \\ \bar{\varkappa}_i &= \lim_{a \rightarrow \infty} \mathbf{E}_i [Z_i(\nu_i(a)) - a], \quad C_i = \mathbf{E}_i \left\{ \log \left(1 + \sum_{k=1}^{\infty} e^{-Z_i(k)} \right) \right\}, \\ I_i &= \int \log \left(\frac{p_i(x)}{p_0(x)} \right) p_i(x) \mu(dx), \quad i = 1, \dots, N. \end{aligned}$$

Note that I_i is the Kullback-Leibler (K-L) information number between the densities $p_i(x)$ and $p_0(x)$. In the i.i.d. case, this number plays the role of the number J_i that appeared in (17.2.5)–(17.2.7).

We will impose the following mild condition on the K-L information numbers I_i :

$$0 < I_i < \infty \quad \text{for } i = 1, \dots, N. \quad (17.2.8)$$

The above condition (finiteness) implies that $\mathbf{E}_i \exp\{-Z_i(1)\} = 1$, and hence, all moments of the negative part $Z_i^-(1) = -\min\{0, Z_i(1)\}$ of LLR's are finite, that is, $\mathbf{E}_i\{Z_i^-(1)\}^m < \infty$ for all $m > 0$. The latter property turns out to be crucial in establishing finiteness of moments of the stopping time τ^* and its asymptotic optimality with respect to moments of the detection delay. Note also that I_i 's are positive whenever the p.d.f.'s $p_0(x)$ and $p_i(x)$ do not coincide almost everywhere, that is when $\mu\{x : p_0(x) \neq p_i(x)\} > 0$.

The following theorem, whose proof is given in the Appendix, establishes higher order asymptotic approximations to the average detection delay of the procedure τ^* and also its first-order asymptotic optimality.

Theorem 17.2.2 *Suppose that condition (17.2.8) holds.*

(i) *Then $A = T$ implies that $\mathbf{E}_\infty(\tau^*) \geq T$, and moreover, $\lim_{T \rightarrow \infty} [\mathbf{E}_\infty \tau^*(T)/T]$ is bounded.*

(ii) *If $A = T$, then for all $m \geq 1$ and $i = 1, \dots, N$*

$$\inf_{\tau \in \mathbf{\Delta}(T)} D_i^m(\tau) \sim D_i^m(\tau^*) \sim \left(\frac{\log T}{I_i} \right)^m \quad \text{as } T \rightarrow \infty. \quad (17.2.9)$$

(iii) *Also assume that $Z_i(1)$ are non-arithmetic and $\mathbf{E}_i |Z_i(1)|^2 < \infty$. Then, as $A \rightarrow \infty$,*

$$\frac{1}{\text{FAR}} = \mathbf{E}_\infty \tau^*(A) = \frac{A}{\bar{\gamma}_N} (1 + o(1)) \quad (17.2.10)$$

and

$$\mathbf{E}_i \tau^*(A) = \frac{1}{I_i} [\log(NA) + \bar{\varkappa}_i - C_i] + o(1), \quad (17.2.11)$$

where $o(1) \rightarrow 0$ as $A \rightarrow \infty$.

It follows that the detection procedure considered minimizes any positive moment of the detection delay for asymptotically small FAR ($T \rightarrow \infty$) whenever the K-L information numbers are finite. Note that finiteness of higher order moments of the LLR is not required. More importantly, if $A = T\bar{\gamma}_N$, then $\mathbf{E}_\infty(\tau^*) \sim T$, which allows us to get an almost exact constraint on the mean time to false alarm (or, equivalently, the FAR) rather than the conservative inequality (17.2.4). Furthermore, we can use (17.2.11) to obtain an asymptotically exact expression for the ADD defined in (17.2.2).

The constants $\bar{\alpha}_i$ and γ_i come from renewal theory (see, for example, Siegmund (1985) and Woodroffe (1982)), and can be computed quite easily in many cases using the technique described in those two books as well as Tartakovsky (1991). The following formulas are particularly useful:

$$\begin{aligned} \gamma_i &= \frac{1}{I_i} \exp \left\{ - \sum_{n=1}^{\infty} n^{-1} [\mathbf{P}_\infty(Z_i(n) > 0) + \mathbf{P}_i(Z_i(n) \leq 0)] \right\}, \\ \bar{\alpha}_i &= \frac{I_i^2 + \sigma_i^2}{2I_i} + \sum_{n=1}^{\infty} \frac{1}{n} \mathbf{E}_i[\min\{0, Z_i(n)\}], \end{aligned} \tag{17.2.12}$$

where $\sigma_i^2 = \mathbf{E}_i[Z_i(1) - I_i]^2$.

The constant C_i is not as straightforward to compute, but Monte Carlo techniques can be used to estimate it accurately in specific examples. Experimentation indicates that formulas (17.2.10) and (17.2.11), with γ_i and $\bar{\alpha}_i$ coming from (17.2.12), give quite accurate approximations for the average run lengths even for moderate values of A .

Two other attractive candidates are the bank of CUSUM (Page's) tests and the bank of Shiryaev-Roberts tests (parallel implementation of CUSUM statistics and Shiryaev-Roberts statistics in channels) that are defined as follows:

$$\begin{aligned} \tau_h &= \min \left\{ n \geq 1 : \max_{1 \leq i \leq N} U_i(n) \geq h \right\}, \\ \tilde{\tau}_{\tilde{A}} &= \min \left\{ n \geq 1 : \max_{1 \leq i \leq N} R_i(n) \geq \tilde{A} \right\}, \end{aligned} \tag{17.2.13}$$

where $U_i(n) = \max_{1 \leq \lambda \leq n} Z_i^\lambda(n)$ is the CUSUM statistic in the i^{th} channel. These procedures have been studied thoroughly in Tartakovsky (1988,

1991a,1991b,1992,1994), for i.i.d. models. It can be shown that $h = \log NT$ implies $\mathbf{E}_\infty(\tau_h) \geq T$, and that $\tilde{A} = NT$ implies $\mathbf{E}_\infty(\tilde{\tau}_{\tilde{A}}) \geq T$. Theorems 17.2.1 and 17.2.2 hold for τ_h and $\tilde{\tau}_{\tilde{A}}$ with corresponding modifications of the thresholds (for example, for the multi-channel CUSUM test, A is replaced with e^h). However, both procedures have a significant drawback compared to τ^* in the sense that there is no analog of (17.2.10) for these procedures. In fact, it is difficult to choose the thresholds h and \tilde{A} so as to guarantee the approximate equalities $\mathbf{E}_\infty(\tau_h) \approx T$ and $\mathbf{E}_\infty(\tilde{\tau}_{\tilde{A}}) \approx T$.

17.2.4 Composite Hypotheses: Adaptive Detection Procedures

So far we have considered the case where both pre-change and post-change distributions were completely known. In practice, a more realistic situation is when the post-change distribution is known only partially — up to parameters θ . Assume that the observations follow a general model with a completely known baseline pdf $p_{0,k}(X_{i,k} | \mathbf{X}_i^{k-1})$ but with $p_{i,k}(X_{i,k} | \mathbf{X}_i^{k-1}) = p_{i,k}(X_{i,k} | \mathbf{X}_i^{k-1}, \theta_i)$ being parameterized by $\theta_i \in \Theta_i$. Write $\mathbf{P}_{\lambda,i,\theta_i}$ and $\mathbf{E}_{\lambda,i,\theta_i}$ for the probability measure and the expectation respectively, when the change occurs in the i^{th} channel at the time-point λ , and the parameter after the change is θ_i .

Let $\hat{\theta}_{i,n} = \hat{\theta}_i(\mathbf{X}_i^n)$ be an estimator of θ_i based on the n observations \mathbf{X}_i^n from the i^{th} channel. For any $k \geq 1$, define the adaptive versions of the partial LLRs

$$\hat{Z}_i^k(k) = \log \frac{p_{i,k}(X_{i,k} | \mathbf{X}_i^{k-1}, \hat{\theta}_{i,k-1})}{p_{0,k}(X_{i,k} | \mathbf{X}_i^{k-1})},$$

which are obtained by replacing the unknown values of the parameters θ_i 's in the LLR's $Z_i^k(k, \theta_i)$ with their estimates based on the previous $k-1$ observations. We stress that the k^{th} observation is not included in the estimate. Similarly, for $n \geq 1$, we define the adaptive versions of the statistics $\hat{R}_i(n)$ and $\hat{R}(n)$,

$$\hat{R}_i(n) = \sum_{\lambda=1}^n \exp \left\{ \sum_{k=\lambda}^n \hat{Z}_i^k(k) \right\}, \quad \hat{R}(n) = N^{-1} \sum_{i=1}^N \hat{R}_i(n). \quad (17.2.14)$$

The adaptive (“plug-in”) detection procedure is defined as

$$\hat{\tau}(A) = \min\{n \geq 1 : \hat{R}(n) \geq A\}. \quad (17.2.15)$$

The first important fact to note is that by replacing $R(n)$ with $\widehat{R}(n)$ in the proof of Lemma 17.2.1 and observing that $\widehat{R}(n) - n$ is a zero-mean martingale (with respect to \mathbf{P}_∞), we immediately obtain: $\mathbf{E}_\infty \widehat{\tau}(A) \geq A$. It follows that $A = T$ implies $\mathbf{E}_\infty \widehat{\tau}(A) \geq T$, and hence, the FAR is easily controlled.

Furthermore, if the quantities $n^{-1} \sum_{k=\lambda}^{n+\lambda-1} \widehat{Z}_i^k(k)$ converge r -quickly to the same positive numbers J_i 's that appeared in Subsection 17.2.3, then an analog of Theorem 17.2.1 holds for the procedure $\widehat{\tau}$. For i.i.d. data models, under certain conditions on the estimators $\widehat{\theta}_{i,n}$, an analog of Theorem 17.2.2 holds as well. Specifically, as $A \rightarrow \infty$, for all $\theta_i \in \Theta_i$ and $i = 1, \dots, N$

$$\mathbf{E}_{1,i,\theta_i} \widehat{\tau}(A) = \frac{1}{I_i} \left[\log(NA) + \frac{1}{2} \log \log(NA) + \widehat{C}_i \right] + o(1), \quad (17.2.16)$$

where \widehat{C}_i is a constant. Comparing (17.2.11) with (17.2.16), we can see that an additional term appears which goes to infinity at the rate of double log of the threshold. This is an unavoidable penalty for the prior uncertainty with respect to the post-change parameter value. We also argue that the rate $(1/2) \log \log T$ cannot be improved, thus implying that the proposed adaptive procedure is asymptotically second-order optimal with respect to the average detection delay $D_{i,\theta_i}(\tau) = \sup_\lambda \mathbf{E}_{\lambda,i,\theta_i}(\tau - \lambda | \tau \geq \lambda)$ in the worst case scenario.

Choosing the estimators $\widehat{\theta}_{i,n}$ is not a straightforward task. For example, if one attempts to detect a change in the common shift parameter $\theta_i = \theta > 0$ of the i.i.d. Gaussian sequence, then the estimate $\widehat{\theta}_{i,n} = n^{-1} S_{i,n}^+$, where $S_{i,n} = \sum_{k=1}^n X_{i,k}$, is not a good choice. This estimate works well only when the change occurs from the very beginning. For large λ , the performance degrades dramatically (Dragalin (1996), Tartakovsky (2003)). A good choice would be an estimator that “forgets” the past. For example, the adaptive exponentially weighted estimators perform fairly well (Dragalin (1996)). To be more specific, for $k = 0, 1, 2, \dots$, define $\widehat{\theta}_{i,k}$ by the recursion

$$\widehat{\theta}_{i,k+1} = \begin{cases} \frac{\beta_k \widehat{\theta}_{i,k} + X_{k+1,i}}{1 + \beta_k} & \text{if } \beta_k \widehat{\theta}_{k,i} + X_{k+1,i} > 0 \text{ and } S_{k+1,i} > 0 \\ \theta_{i,0} & \text{otherwise} \end{cases} \quad (17.2.17)$$

with the initial condition $\widehat{\theta}_{i,0} = \theta_{i,0}$, where $\theta_{i,0}$ is a design positive number and $\beta_k, k = 0, 1, \dots$ satisfy the recursion

$$\beta_{k+1} = (1 + \beta_k) \mathbb{1}_{\{\beta_k \widehat{\theta}_{k,i} + X_{k+1,i} > 0, S_{k+1,i} > 0\}}, \quad \beta_0 = 0.$$

Hereafter $\mathbb{1}_\Omega$ denotes an indicator of a set Ω . Therefore, the adaptive exponentially weighted estimate (17.2.17) has a structure similar to a sample mean where the current sample size k is replaced with the “adaptive” number β_k , which is set to zero (renewed) whenever the statistic $S_{k,i}$ hits the zero level. This allows us to forget the observations that are not consistent with the “change” hypothesis. A reasonable choice for the initial condition $\theta_{i,0}$ is the minimum expected value of the change in the mean.

17.3 DETECTION IN DISTRIBUTED SENSOR SYSTEMS

Distributed sensor systems capable of collecting, storing, and disseminating a variety of environmental data have the potential to enable the next revolution in information technology. An important application area for distributed sensors is in *environmental awareness* systems. Examples of applications include toxic agent detection, intruder detection for homes and businesses, child-care monitoring systems, automated airport surveillance and metal detection, detection of the onset of an epidemic, failure detection in manufacturing systems and large machines, and patient monitoring in hospitals and homes. A key component of these environmental awareness systems is change detection based on observations made by the sensors.

As described in Section 17.1, the distributed sensor problem could in general be a multichannel change-point detection problem. However, for simplifying the presentation, we assume a single-channel version of the problem, with the understanding that the multichannel generalization is straightforward.

Suppose there are L sensors in the system. At time n , an observation $X_{\ell,n}$ is made at sensor S_ℓ . Conditioned on the change-point λ , the observation sequences $\{X_{1,n}\}, \{X_{2,n}\}, \dots, \{X_{L,n}\}$ are assumed to be mutually independent. Furthermore, throughout this section, we restrict our attention to the i.i.d. case where the observations in a particular sequence, say $\{X_{\ell,n}\}_{n \geq 1}$, are independent conditionally on λ , have a common p.d.f. $f_\ell^{(0)}$ before the change, and a common p.d.f. $f_\ell^{(1)}$ from the time of change. Note that we are assuming that all the sensors change distribution at the change-time λ . Let \mathbf{P}_λ (\mathbf{P}_∞) and \mathbf{E}_λ (\mathbf{E}_∞) stand for the probability measure and the corresponding expectation

when the change occurs at time λ (does not occur, that is $\lambda = \infty$).

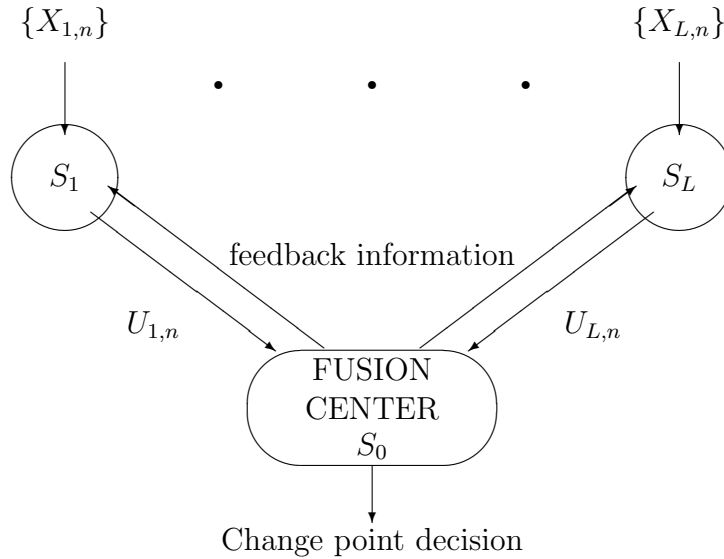


Figure 17.3.1: Block-Diagram of a Decentralized Multi-Sensor Change-Point Detection System.

A general block-diagram of the decentralized change-point detection system is shown in Figure 17.3.1. Based on the information available at S_ℓ at time n , a message $U_{\ell,n}$, belonging to a finite alphabet of size V_ℓ , is formed and sent to the fusion center. We will use the vector notation: $\mathbf{X}_n = (X_{1,n}, \dots, X_{L,n})$ and $\mathbf{U}_n = (U_{1,n}, \dots, U_{L,n})$. Based on the sequence of sensor messages, a decision about the change is made at the fusion center. The fusion center picks a time τ which is a *stopping time* on $\{\mathbf{U}_n\}_{n \geq 1}$, and a change is declared to have occurred at *tau*.

Various information structures are possible for the decentralized configuration depending on how feedback and local information are used at the sensors (Veeravalli (2001)). Consider the simplest information structure where the message $U_{\ell,n}$ formed by sensor S_ℓ at time n is a function of only its current observation $X_{\ell,n}$, i.e., $U_{n,\ell} = \psi_{\ell,n}(X_{\ell,n})$. Moreover, since for a particular ℓ , the sequence $\{X_{\ell,n}\}_{n \geq 1}$ is assumed to be i.i.d., it is natural to confine ourselves to *stationary* quantizers for which the quantizing functions $\psi_{\ell,n}$ do not depend on n , i.e. $\psi_{\ell,n} = \psi_\ell$ for all $n \geq 1$.

The quantizing functions $\{\psi_\ell; \ell = 1, \dots, L\} = \boldsymbol{\psi}$, together with the fusion center stopping time τ , form a policy $\phi = (\tau, \boldsymbol{\psi})$. The goal is to

choose the policy ϕ that minimizes $D^m(\phi)$ defined by

$$D^m(\phi) = \sup_{1 \leq \lambda < \infty} \mathbf{E}_\lambda \{ (\tau - \lambda)^m \mid \tau \geq \lambda \} \quad (17.3.1)$$

for all $m > 0$, while maintaining the average time to false alarm, $\mathbf{E}_\infty(\tau)$, at a level greater than T .

Let H_λ be the hypothesis that the change occurs at time $\lambda \in \{1, 2, \dots\}$, and let H_∞ be the hypothesis that $\lambda = \infty$ (no change). Since the observations at each sensor S_ℓ , $\{X_{\ell,n}, n = 1, 2, \dots\}$, are i.i.d., the sensor outputs, $\{U_{\ell,n}, n = 1, 2, \dots\}$ will also be i.i.d. for stationary sensor quantizers. Let $g_\ell^{(j)}$ denote the distribution induced on $U_{\ell,n}$ when the observation $X_{\ell,n}$ is distributed as $f_\ell^{(j)}$, $j = 0, 1$.

Then, for *fixed* sensor quantizers, the LLR between the hypotheses H_λ and H_∞ at the sensor S_ℓ and at the fusion center are respectively given by

$$Z_\ell^\lambda(n) = \sum_{k=\lambda}^n \log \frac{g_\ell^{(1)}(U_{\ell,k})}{g_\ell^{(0)}(U_{\ell,k})} \quad \text{and} \quad Z^\lambda(n) = \sum_{\ell=1}^L Z_\ell^\lambda(n). \quad (17.3.2)$$

For fixed sensor quantizers, the fusion center faces a standard change detection problem based on the vector observation sequence $\{\mathbf{U}_k\}$. Hence we can define the Shiryaev-Roberts statistic $R(n)$ that obeys the recursion:

$$R(n) = (1 + R(n-1))e^{Z^n(n)}, \quad R(0) = 0. \quad (17.3.3)$$

Then the Shiryaev-Roberts detection procedure at the fusion center $\tau^* = \tau^*(A)$ is given by

$$\tau^*(A) = \min \{n \geq 1 : R(n) \geq A\} \quad (\tau^* = \infty \text{ if no such } n \text{ exists}), \quad (17.3.4)$$

where A is a positive threshold which is selected so that $\text{FAR}(\tau^*(A)) \leq 1/T$.

Let $I(g_\ell^{(1)}, g_\ell^{(0)})$ denote the K-L information number between the densities $g_\ell^{(1)}$ and $g_\ell^{(0)}$. Assume that

$$0 < \sum_{\ell=1}^L I(g_\ell^{(1)}, g_\ell^{(0)}) < \infty. \quad (17.3.5)$$

Then an application of Theorem 17.2.2 (ii) (with one channel) gives us that the detection procedure $\tau^*(A)$ defined in (17.3.4), with $A = T$, is

asymptotically optimal as $T \rightarrow \infty$ among all procedures with FAR no greater than $1/T$. To be specific, if $A = T$, then for all $m > 0$

$$\inf_{\tau \in \mathbf{\Delta}(T)} D^m(\tau) \sim D^m(\tau^*) \sim \left(\frac{\log T}{\sum_{\ell=1}^L I(g_\ell^{(1)}, g_\ell^{(0)})} \right)^m \quad \text{as } T \rightarrow \infty.$$

This result immediately reveals how to choose the sensor quantizers.

Corollary 17.3.1 *It is asymptotically optimum (as $T \rightarrow \infty$) for sensor S_ℓ to pick ψ_ℓ to maximize the K-L information number $I(g_\ell^{(1)}, g_\ell^{(0)})$.*

Based on the results of Tsitsiklis (1993), it is easy to show that the optimal stationary quantizer $\psi_{\ell,\text{opt}}$ is a *monotone likelihood ratio quantizer* (MLRQ), that is, there exist thresholds $\beta_{\ell,1}, \beta_{\ell,2}, \dots, \beta_{\ell,V_\ell-1}$ satisfying $0 = \beta_{\ell,0} \leq \beta_{\ell,1} \leq \beta_{\ell,2} \leq \dots \leq \beta_{\ell,V_\ell-1} < \infty = \beta_{\ell,V_\ell}$ such that

$$\psi_{\ell,\text{opt}}(X) = i \quad \text{only if} \quad \beta_{\ell,i-1} < \frac{f_\ell^{(1)}(X)}{f_\ell^{(0)}(X)} \leq \beta_{\ell,i}, \quad i = 1, \dots, V_\ell.$$

Thus, the asymptotically optimal policy ϕ_{opt} for a decentralized change detection problem consists of a stationary (in time) set of MLRQ's at the sensors followed by a Shiryaev-Roberts procedure based on $\{\mathbf{U}_k\}$ at the fusion center (as described in (17.3.4)). In other words, if we denote by τ_{opt}^* the Shiryaev-Roberts stopping rule at the fusion center for the case where the sensor quantizers are chosen to be $\boldsymbol{\psi}_{\text{opt}} = \{\psi_{\ell,\text{opt}}\}$, then the asymptotically optimal policy $\phi_{\text{opt}} = (\tau_{\text{opt}}^*, \boldsymbol{\psi}_{\text{opt}})$.

For each ℓ , let the p.d.f.'s induced on $U_{\ell,m}$ by the optimal MLRQ $\psi_{\ell,\text{opt}}$ be given by $g_{\ell,\text{opt}}^{(1)}$ and $g_{\ell,\text{opt}}^{(0)}$. Then the effective K-L information number between the "change" and "no change" hypotheses at the fusion center is given by

$$I_{\text{tot}} = \sum_{\ell=1}^L I(g_{\ell,\text{opt}}^{(1)}, g_{\ell,\text{opt}}^{(0)}).$$

Finally, we denote by $\Phi_{\text{st}}(T)$ the class of policies ϕ with all stationary quantizers and stopping rules at the fusion center such that $\tau \in \mathbf{\Delta}(T)$.

The asymptotic performance of the asymptotically optimal solution to the decentralized change detection problem described above is given in the following theorem, which follows directly from Theorem 17.2.2.

Theorem 17.3.1 *Suppose $0 < I_{\text{tot}} < \infty$.*

(i) *Then $A = T$ implies that $\mathbf{E}_{\infty} \tau_{\text{opt}}^* \geq T$, and moreover, $\lim_{T \rightarrow \infty} [\mathbf{E}_{\infty} \tau_{\text{opt}}^*(T)/T]$ is bounded.*

(ii) *If $A = T$, then for all $m \geq 1$,*

$$\inf_{\phi \in \Phi_{\text{st}}(T)} D^m(\phi) \sim D^m(\phi_{\text{opt}}) \sim \left(\frac{\log T}{I_{\text{tot}}} \right)^m \quad \text{as } T \rightarrow \infty.$$

(iii) *In addition, assume that $Z^1(1)$ is non-arithmetic and that the second*

moment $\mathbf{E}_1 |Z^1(1)|^2$ is finite. Then, as $A \rightarrow \infty$,

$$\text{FAR}(\tau^*(A)) = \frac{1}{\mathbf{E}_{\infty} \tau_{\text{opt}}^*(A)} = \frac{\gamma}{A} (1 + o(1)), \quad (17.3.6)$$

$$\text{ADD}(\phi_{\text{opt}}) = \mathbf{E}_1 \tau_{\text{opt}}^*(A) - 1 = \frac{1}{I_{\text{tot}}} (\log A + \bar{\varkappa} - C) - 1 + o(1). \quad (17.3.7)$$

The quantities γ , $\bar{\varkappa}$, and C are defined in a similar manner as γ_i , $\bar{\varkappa}_i$, and C_i in Theorem 17.2.2 if we replace $Z_i(n)$ by $Z(n) = \sum_{\ell=1}^L Z_{\ell}^1(n)$.

The above result not only allows us to choose the threshold A to meet the FAR constraint precisely, but also to calculate the corresponding ADD.

Remark 17.3.1: The condition that the LLR $Z^1(1)$ be non-arithmetic is imposed because one needs to consider certain discrete cases separately in the renewal theorem (Woodroffe (1982), Section 2.1). Since the data at the output of quantizers are discrete, it may happen that the LLR does not obey this condition. If $Z^1(1)$ is arithmetic with span $d > 0$, the results of Theorem 17.3.1 (iii) hold true as $\log A \rightarrow \infty$ through multiples of d (that is, $\log A = kd$, $k \rightarrow \infty$) and with respective modification of definition of \varkappa . However, even in most discrete cases the LLR remains non-arithmetic (see, for example, Section 17.4.3).

In Subsection 17.4.3, we will give an example of target detection where the sensor observations are Gaussian random variables with different statistics before and after the change. We will verify the asymptotic results given above via Monte-Carlo simulations.

17.4 APPLICATIONS AND EXPERIMENTAL RESULTS

17.4.1 Target Detection and Tracking in Surveillance Systems

Surveillance systems, such as those in ballistic and cruise missile defense, deal with the detection and tracking of moving targets. Infrared Search and Track (IRST) systems are one component of a multisensor suite which can meet the technical challenge of the timely detection/track/identification of small targets. The most challenging problem for an IRST system is the rapid detection of a maneuvering dim target against a heavily cluttered background. To illustrate the importance of this task, we remark that under certain conditions *a few seconds' decrease* in the time it takes to detect a sea/surface skimming cruise missile can lead to a significant increase in the *probability of raid annihilation*.

The problem of detecting moving targets is complicated by the fact that target tracks occur (appear and disappear) at unknown points in time. As a result, this problem can be naturally formulated as an abrupt change detection problem, which is the central theme of this chapter.

Assume that there is a mosaic of staring IR sensors that register the data $\mathbf{X}_n = \{X_n(r_{ij}), i, j = 1, \dots, M\}$, $n \geq 1$, which represent a sequence of 2D frames of intensities $X_n(r_{ij})$ at the pixels with coordinates r_{ij} . Typically, the observations are highly cluttered. Since clutter is usually hundreds or even thousands of times more intensive than sensor noise and the intensity of the target, the detection is impossible without clutter suppression. For this reason, we first apply a spatial-temporal clutter rejection and electronic scene stabilization algorithm (filter) which was developed by Tartakovsky and Blažek (2000). At the output of this filter, we observe the residuals

$$X_n(r) = \mathbb{1}_{\{\lambda \leq n\}} \theta S(r - y_n) + \xi_n(r),$$

where λ is an unknown moment of target appearance; $\mathbb{1}_\Omega$ is an indicator of the set Ω ; θ is an unknown intensity of the target; $S(r)$ is a known *point spread function* (PSF) of the sensor; y_n is an unknown spatial location of a target on the plane at time n ; $\xi_n(r)$ is the effective noise (residual clutter plus noise) having variance σ^2 . It is worth mentioning

that the PSF $S(r)$ has a very compact support—its effective spatial size is normally 2×2 or 3×3 pixels. Also, even after clutter removal, pixel SNR $q_0 = \theta^2/\sigma^2$ can be as small as 0 dB or even negative.¹

The detection is complicated by the fact that the velocities and the positions of targets are unknown. In order to overcome this difficulty, we use a so-called “*track-before-detect*” (TBD) approach (Kligys et al. (1998), Petrov et al. (submitted), Rozovskii and Petrov (1999)) which is embedded into the multichannel system that represents a bank of position and velocity filters. The idea of TBD is to perform “preliminary” tracking in an attempt to align successive frames according to typical patterns of the targets’ dynamics. The second block in the developed system implements this idea. Specifically, we use switching multiple models for target dynamics y_n and a *bank of optimal nonlinear filters* (BONF) to estimate the position of the target at each point in time. The output of BONF is the unnormalized posterior density for y_n given previous data. Maximizing this density, we obtain the estimate \hat{y}_n of the target location y_n . This estimate is then fed into the detection block along with the estimate of the target intensity $\hat{\theta}_n$. The BONF-based TBD algorithm has been developed by B. Rozovsky and A. Petrov, and its detailed description can be found in Petrov et al. (submitted) and Rozovskii and Petrov (1999).

Now we are in a position to describe the detection algorithm. The algorithm has the form (17.2.14) and (17.2.15) where the statistic $\hat{R}_i(n)$ corresponds to the i^{th} position (spatial) channel in the bank of matched filters. Note that we need a bank of filters, since the estimates \hat{y}_n of the target position are not perfect. As a result, there is almost always a mismatch between the actual target position and its estimated value. The required number of filters in the bank is, however, relatively small compared to the case where TBD is not performed. For example, we used $N = 9 - 32$ filters in the bank regardless of the size of the image. (Compare to $N = 2^{12}$ which would be required for the system without TBD comprised of the bank of 3D matched filters for the image size 128×128 and the effective target size 2×2 pixels.) Equivalent noise is modeled by the zero-mean white Gaussian process, in which case the partial LLR for the i^{th} position channel is defined by equation (17.4.1), with $\hat{\theta}_k$ being an estimate of the target intensity and δ_i being a two-dimensional shift that is measured in the number of pixels ($i = 0, \dots, N - 1$). Here, the estimate $\hat{\theta}_k$ has been computed by using

¹SNR in decibels (dB) is defined as $10 \log_{10}(\theta^2/\sigma^2)$.

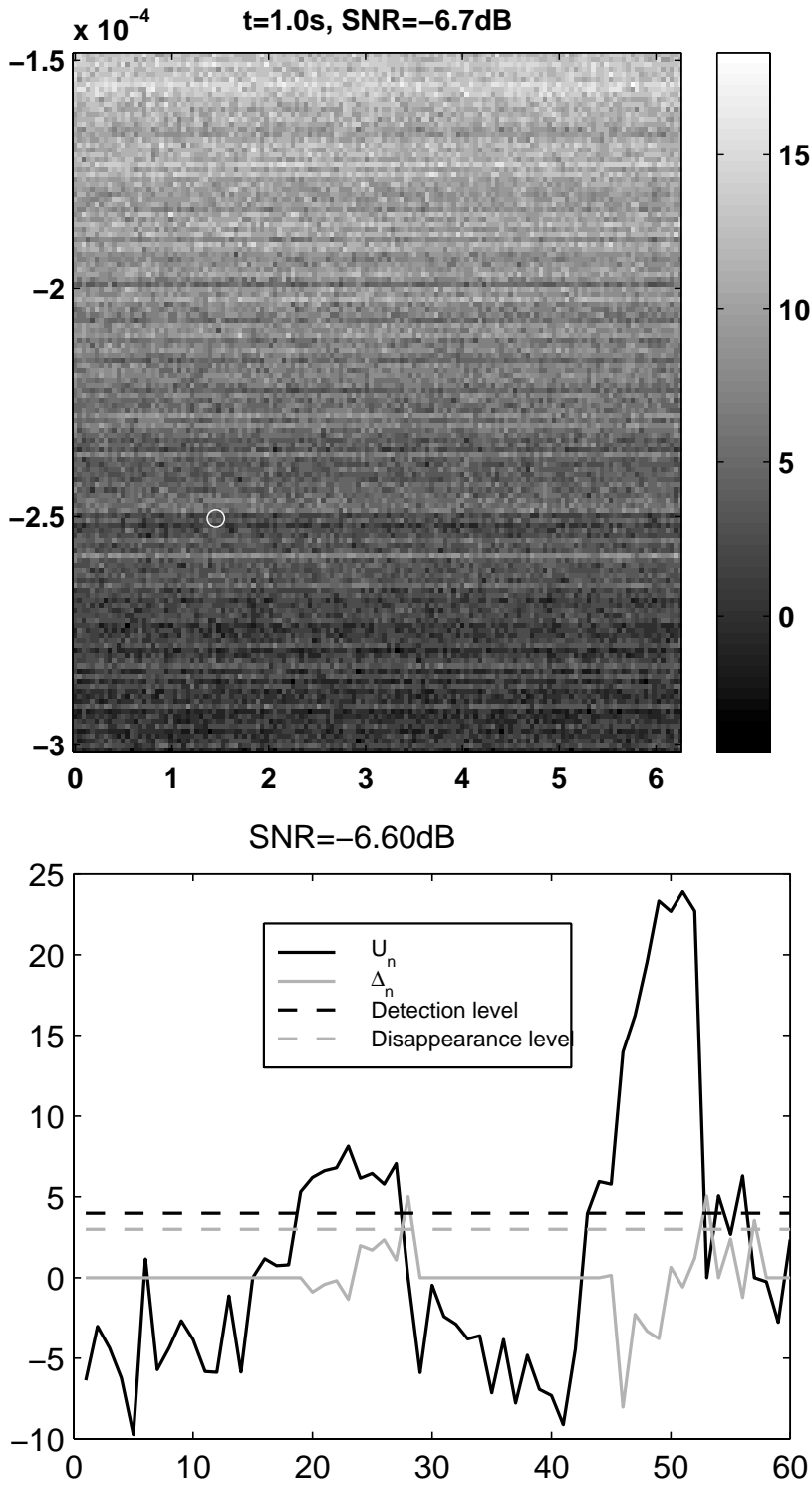


Figure 17.4.1: Detection of Track Appearance and Disappearance: Real IR Background at the Output of the De-Cluttering Filter (top) and Target Detection (bottom).

a formula similar to (17.2.17).

$$\widehat{Z}_i^k(k) = \frac{\widehat{\theta}_{k-1}}{\sigma^2} \sum_{l,j} S(r_{lj} - \widehat{y}_k - \delta_i) X_k(r_{lj}) - \frac{\widehat{\theta}_{k-1}^2}{2\sigma^2} \sum_{l,j} S^2(r_{lj} - \widehat{y}_k - \delta_i), \quad (17.4.1)$$

Thus, the results of Subsection 17.2.4 are applied by using the LLR's (17.4.1). To be precise, detections are declared and tracks are initiated when the statistic $\widehat{R}(n)$ exceeds the threshold A . The threshold is chosen so that the average frequency of false exceedances would be bounded above by a given level $\text{FAR} \leq \bar{F} = 1/T$. According to Lemma 17.2.1, $A = 1/\bar{F}$ guarantees this inequality.

Figure 17.4.1 illustrates the results of detection by the described adaptive change-point detection algorithm. In these experiments, we used the real cluttered and jittery IR background obtained from the NAVY SPAWAR Systems Center, San Diego, CA (staring shipboard IRST, LAPTEX field test). The picture on the left-hand side shows a typical data frame at the output of the clutter rejection filter. The effective size of the target is 3×3 pixels, and it is invisible to the naked eye, since the pixel SNR varies from frame to frame in the range -16 to 0 dB. For this reason, it is explicitly circled in the picture. The picture on the right side depicts two statistics: $U_n(\widehat{y}_n) = \log \widehat{R}(n)$ and Δ_n . the first one is sensitive to target appearance and the second one to target disappearance. The detection of tracks occurs when the adaptive statistic $U_n(\widehat{y}_n)$ exceeds the threshold $\log A$ (the upper one) and track disappearance is declared when the statistic Δ_n exceeds the threshold C (the lower one). See Petrov et al. (submitted) and Tartakovsky et al. (1999) for more details. The frame rate was 1 frame/second and the threshold A was chosen so as to guarantee that there would be no more than 1 false detection per minute (that is, per 60 frames). In the particular scenario shown in the figure, the algorithm detected the first target with the delay of about 20 seconds (20 frames). This number matches the first-order approximation given by Theorem 17.2.1 of Subsection 17.2.4 (see (17.2.6)),

$$\mathbf{E}_{i,\lambda,\theta}(\tau^* - \lambda | \tau^* \geq \lambda) \approx (2/q) \log(N/\bar{F}),$$

where $q = \frac{\theta^2}{\sigma^2} \sum_{l,j} S^2(r_{lj})$ is the cumulative SNR and N is the number of spatial (position) channels. Indeed, for $N = 9$ channels used in the experiments, $\bar{F} = 1/60$, and estimated cumulative SNR -1.14 dB, the average detection delay is $\text{ADD} \approx 16.4$ according to this approximate formula.

The results obtained allow us to conclude that the adaptive algorithm developed above is able to detect even very low SNR targets with reasonable detection delays, and that the theoretical results of Section 17.2 are useful for performance evaluation.

17.4.2 Rapid Attack/Intrusion Detection

In computer networks, large-scale attacks in their final stages can readily be identified by observing very abrupt changes in the network traffic. In the early stage of an attack, however, these changes are hard to detect and difficult to distinguish from usual traffic fluctuations. Existing intrusion detection systems can be classified as either *Signature Detection Systems* or *Anomaly Detection Systems*. The latter systems compare the parameters of the observed traffic with “normal” (legitimate) network traffic. The attack is declared once a deviation from normal traffic is observed (Kent (2000)).

Our approach belongs to the class of Anomaly Detection Systems. The idea is based on the observation that an attack leads to relatively abrupt changes in the statistics of the traffic when compared to the traffic’s “normal mode.” Therefore, the problem of detecting an attack can be formulated and solved as a change-point detection problem: detect a change in the traffic model with minimal average delay, while controlling FAR. In this subsection, we briefly describe an efficient adaptive sequential method, recently developed in Blažek et al. (2001,2003), for an early detection of intrusions from the class of “*Denial-of-Service* (DoS)” attacks.

In experiments, we used a Network Simulator NS-2² with a network consisting of 100 nodes configured into a transit-stub topology that is depicted in Figure 17.4.2. The network contained one transit domain, four transit nodes, and 12 stub domains with 96 nodes. Further details can be found in Blažek et al. (2001,2003).

While monitoring network traffic, one can observe various kinds of features related to the headers and sizes of the received and transmitted packets, the usage of system resources, service quality, and similar aspects associated with the utilization of the network. For example, in the transport layer, we observe the number of TCP (*Transmission Control Protocol*) packets categorized by size and/or type, the number of UDP (*User Datagram Protocol*) packets and their sizes, and the source

²More information on the NS-2 can be found at <http://www.isi.edu/nsnam/ns/>

and destination port for each packet. An intrusion leads to a change

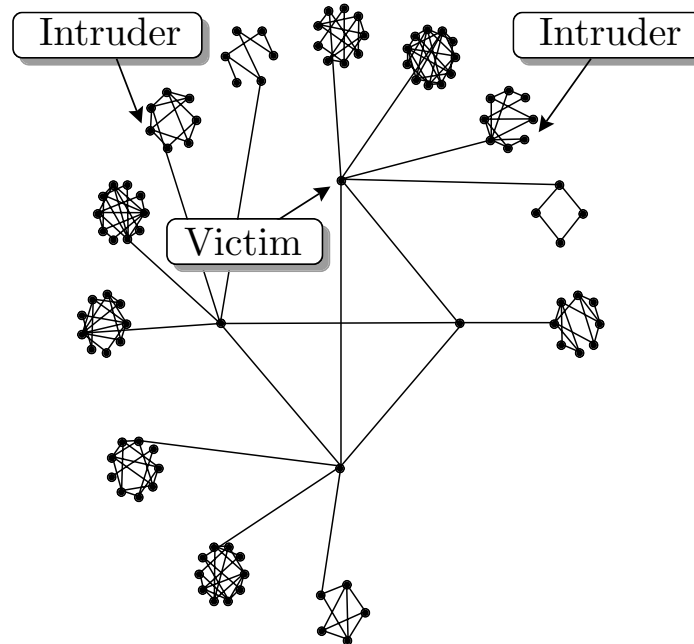


Figure 17.4.2: Transit-Stub Network Topology Used in Experiments

in the traffic intensity through changes in packet sizes. Therefore, the problem of detecting a DoS attack is regarded as the quickest change detection in the mean values of the numbers of packets. Let pt denote the type of the packet—ICMP, UDP, or TCP. In the experiments, we categorized the packets by their sizes into a number of size bins. Let $X_{pt}^{k,i}$ be the total number of packets of type pt with sizes in the i^{th} bin received during the k^{th} time interval, where $i = 1, \dots, M$ (M being the number of bins). Therefore, the total monitoring system is a multichannel system with $N = 3M$ channels.

In network security applications, it is fairly difficult to build an exact statistical model. As a result, the post-change distributions $\mathbf{P}_{pt,i}$'s are usually not specified. For this reason, we used a nonparametric version of the CUSUM-type algorithm (17.2.13). Specifically, we performed simultaneous thresholding of the statistics $S_{k,i}^{pt}$ with reflection from the zero barrier that were functions of the numbers of packets $X_{pt}^{k,i}$ and their “historical” mean values $\mu_{i,k}$. If the statistic $\max_{pt,i}(S_{k,i}^{pt})$ exceeds a threshold h , then an alarm message is sent to the decision making engine. The statistic $S_{k,i}^{pt}$ represents a nonparametric adaptive version

of the CUSUM statistic $U_i(k)$ defined at the end of Subsection 17.2.3.

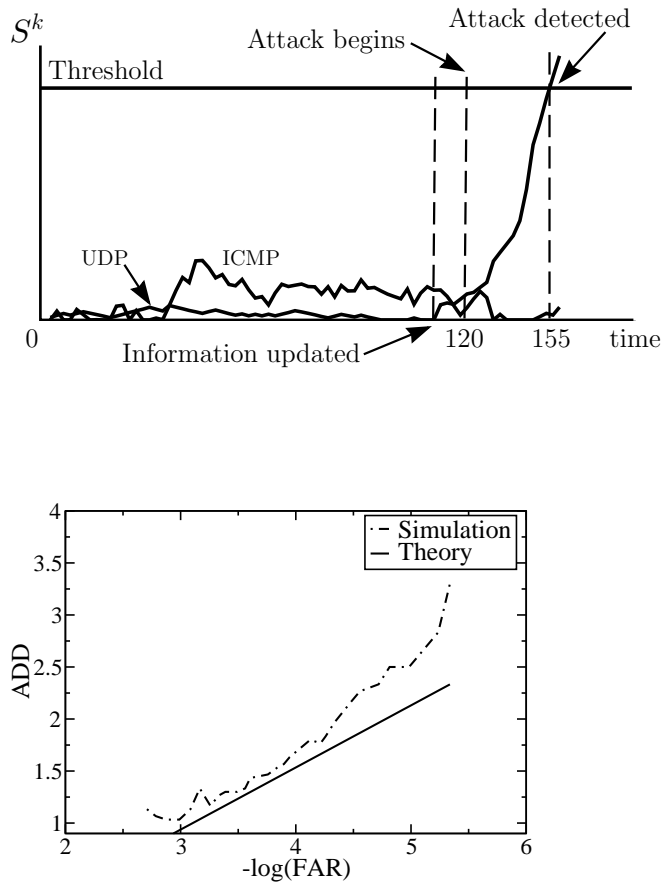


Figure 17.4: Intrusion Detection: One Particular Run of a UDP DoS Attack (Above) and Operating Characteristic of the Detection Algorithm (Below).

As shown in Figure 17.4.3 (top part), information about the patterns of regular data flow is updated when the statistic $S_{k,i}^{pt}$ reaches the zero barrier. If the decision-making engine reports that a previously issued alarm message is a false alarm, then the information about regular data patterns and thresholds is updated accordingly (in particular, the pre-change mean value is re-estimated), and the data monitoring starts all over again.

Under regular conditions, the traffic consisted of approximately 5%

ICMP packets, 20% UDP packets, and 75% TCP packets. After a 120-second period (measured using the simulator time) of regular traffic, we initiated an attack (TCP SYN Flooding, UDP Packet Storm or ICMP Ping Flooding; see Blažek et al. (2001) and CERT. (1996)) and traffic rapidly increased, reaching 20% of all traffic.

As shown in Figure 17.4.3 (top), the sequential algorithm has detected the UDP DoS attack in its early stage. The plot at the bottom illustrates the operating characteristic (ADD versus FAR) obtained by simulations (dashed line) and by the first-order asymptotic formulas similar to (17.2.6) and (17.2.9) (solid line) for the UDP Packet Storm attack. The plot of the theoretical estimate of ADD versus $|\log \text{FAR}|$ is the straight line with the slope that can be computed from the asymptotic theory (see formulas (17.2.6) and (17.2.9)). It is seen that the experimental estimates of ADD are always bigger than the theoretical estimates. This is not surprising, since the first-order approximations used in calculations ignore excesses over the thresholds of the decision statistics. However, the accuracy of these approximations is high enough to be useful for preliminary estimates.

Therefore, as in the previous subsection, the results of the asymptotic theory can be used to predict the performance of the detection algorithm with a reasonable accuracy.

17.4.3 Decentralized Detection Example and Simulation Results

Consider the problem of detecting a slowly fluctuating target using L geographically separated sensors (for example, radars). The observations are corrupted by additive white Gaussian noise (sensor noise or residual clutter and noise) that is independent between each two sensors. The sensors pre-process the observations using a matched filter, matched to the signal corresponding to the target. The output of the matched filter at sensor S_ℓ at time n is given by:

$$X_{\ell,n} = \begin{cases} \xi_{\ell,n} & \text{if } n < \lambda \\ \mu_\ell + \xi_{\ell,n} & \text{if } n \geq \lambda, \end{cases}$$

where λ is the time of appearance of the target, $\{\xi_{\ell,n}\}_{n \geq 1}$ is a sequence of i.i.d. zero-mean Gaussian random variables with variance σ_ℓ^2 , and μ_ℓ is the mean value that is related to the average intensity of the target.

Therefore, the likelihood ratio at sensor S_ℓ is given by

$$Y_\ell(X_{\ell,n}) = \exp \left\{ (X_{\ell,n} \mu_\ell - \mu_\ell^2/2) / \sigma_\ell^2 \right\}, \quad n = 1, 2, \dots$$

Since $Y_\ell(x)$ is monotonically increasing in x , we can characterize the sensor quantizers in terms of thresholds on the observations, rather than on their likelihood ratios. To further simplify the example, we assume that the sensor messages are binary, that is, $V_\ell = 2$ for all ℓ . Then the quantizers reduce to binary tests that are characterized by a single threshold, that is,

$$U_{\ell,k} = \begin{cases} 1 & \text{if } X_{\ell,k} \geq \beta_\ell \\ 0 & \text{if } X_{\ell,k} < \beta_\ell \end{cases}.$$

The distribution induced on $U_{\ell,k}$ by this quantizer is given by:

$$g_\ell^{(j)}(0) = 1 - g_\ell^{(j)}(1) = \Phi \left(\frac{\beta_\ell - j\mu_\ell}{\sigma_\ell} \right) = q_\ell^{(j)}, \quad j = 0, 1, \quad (17.4.2)$$

where $\Phi(\cdot)$ is the distribution function of a standard Gaussian random variable. The optimal value of β_ℓ , i.e., the one that maximizes $I(g_\ell^{(1)}, g_\ell^{(0)})$, is easily found based on (17.4.2). Then we can compute the Shiryaev-Roberts statistic for the fusion center using (17.3.3). In particular, it is easy to show that

$$Z^n(n) = \sum_{\ell=1}^L \left\{ U_{\ell,n} \log \left[\frac{1 - q_{\ell,\text{opt}}^{(1)} q_{\ell,\text{opt}}^{(0)}}{1 - q_{\ell,\text{opt}}^{(0)} q_{\ell,\text{opt}}^{(1)}} \right] - \log \left[\frac{q_{\ell,\text{opt}}^{(0)}}{q_{\ell,\text{opt}}^{(1)}} \right] \right\}, \quad (17.4.3)$$

where $q_{\ell,\text{opt}}^{(j)}$'s are the optimal values that correspond to (17.4.2) with the optimal threshold values $\beta_\ell = \beta_{\ell,\text{opt}} = \text{argmax } I(g_\ell^{(1)}, g_\ell^{(0)})$.

An example with five sensors having identically distributed observations is illustrated in Figure 17.4.4. The parameter values are $\mu_\ell = 0.4$ and $\sigma_\ell^2 = 1$. The K-L information number for the sensor observations is 0.08. The optimal threshold that maximizes the K-L information number at the output of the sensor is $\beta = 0.32$, and the corresponding maximum K-L number after quantization is 0.0509. We plot ADD versus FAR for the optimal decentralized detection policy and compare the performance with that of a centralized policy which has direct access to the observations at the radars. As we expect, for the centralized policy, the plot of ADD versus $-\log(\text{FAR})$ is a straight line with a slope that

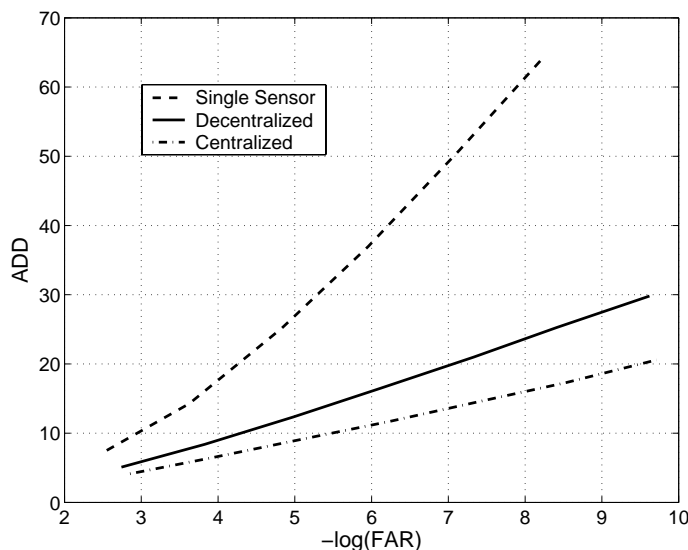


Figure 17.4.4: Operating Characteristic for Five Identically Distributed Sensors.

is approximately equal to $1/[5I(f^{(1)}, f^{(0)})] = 2.5$. For the optimal decentralized policy, the tradeoff curve between ADD and $-\log(\text{FAR})$ has a slope that is approximately equal to $1/I_{\text{tot}} \approx 3.93$, as expected from Theorem 17.3.1. 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 a decentralized setting, especially with binary quantizers. Also shown in Figure 17.4.4 is the tradeoff curve for a centralized detection policy with a *single* sensor. As expected, the slope of ADD versus $-\log(\text{FAR})$ is five times greater than that in the five sensor centralized case. Furthermore, it can be seen that even if the sensor observations are quantized to one bit, the decentralized policy with five sensors far outperforms the single-sensor centralized policy.

Based on (17.4.3), we may also compute higher order approximations for FAR and ADD as given in Theorem 17.3.1 using formulas (17.2.12). These higher-order approximations typically match the simulation results very closely. Results of the detailed analysis will be presented elsewhere.

17.5 CONCLUDING REMARKS

For the problem of detecting abrupt changes in multichannel and distributed systems, we have proposed easily implementable sequential procedures based on multichannel versions of Shiryaev-Roberts and CUSUM-type statistics. A major theoretical result is the asymptotic optimality property of these procedures among all detection procedures with a guaranteed low rate of false alarms for quite general non-i.i.d. models. As we have shown, the regularity conditions can be relaxed considerably for the i.i.d. models. Further, we have derived lower bounds for the mean time to false alarm and asymptotic expansions for the average detection delay. These bounds and asymptotic expansions provide useful approximations for the operating characteristics and constitute the basis for the approximate design of detection thresholds.

The usefulness of the asymptotic theory developed here has been verified by extensive simulations and, more importantly, by implementation using real data. In particular, the proposed detection algorithms have been implemented for rapid detection of moving dim targets against heavy clutter byIRST systems, as well as for real-time intrusion detection in high-speed computer networks. The evaluation of the performance of these algorithms in the existing testbeds illustrates their high efficiency. Based on these results, we can conclude that the technology developed here allows one to reliably detect even low-SNR targets while maintaining the desired frequency of false detections. In addition, it helps detect denial-of-service attacks in their early stages, well before the hostile traffic reaches its full potential.

APPENDIX

Appendix A1. Proof of Theorem 17.2.1

Write

$$M_i(A) = J_i^{-1} \log(NA), \quad \gamma_{\lambda,i}(\varepsilon, A) = \mathbf{P}_{\lambda,i} \{0 \leq \tau^* - \lambda < (1 - \varepsilon)M_i(A)\}.$$

By Chebyshev's inequality, for every $\varepsilon > 0$

$$\mathbf{E}_{\lambda,i}[(\tau^* - \lambda)^m \mathbb{1}_{\{\tau^* \geq \lambda\}}] \geq [\varepsilon M_i(A)]^m \mathbf{P}_{\lambda,i} \{\tau^* - \lambda \geq \varepsilon M_i(A)\}.$$

Obviously,

$$\mathbf{P}_{\lambda,i} \{\tau^* - \lambda \geq (1 - \varepsilon)M_i(A)\} = \mathbf{P}_{\lambda,i} \{\tau^* \geq \lambda\} - \gamma_{\lambda,i}(\varepsilon, A),$$

and $\mathbf{P}_{\lambda,i} \{\tau^* \geq \lambda\} = \mathbf{P}_\infty \{\tau^* \geq \lambda\}$, since the event $\{\tau^* \geq \lambda\}$ belongs to the sigma algebra $\mathcal{F}_{\lambda-1}$. It follows that for every $0 < \varepsilon < 1$

$$\mathbf{E}_{\lambda,i} [(\tau^* - \lambda)^m | \tau^* \geq \lambda] \geq [(1 - \varepsilon)M_i(A)]^m \left[1 - \frac{\gamma_{\lambda,i}(\varepsilon, A)}{\mathbf{P}_\infty \{\tau^* \geq \lambda\}} \right].$$

It can be shown that $\mathbf{P}_\infty \{\tau^* \geq \lambda\} \rightarrow 1$ as $A \rightarrow \infty$ for any fixed λ . Also, applying an argument similar to that used in the proof of Lemma 3.1 in Tartakovsky (2003), we conclude that for an arbitrary $\varepsilon \in (0, 1)$, $\gamma_{\lambda,i}(\varepsilon, A) \rightarrow 0$ as $A \rightarrow \infty$. Since ε is arbitrary, we obtain the following asymptotic lower bound

$$\mathbf{E}_{\lambda,i} [(\tau^* - \lambda)^m | \tau^* \geq \lambda] \geq \left(\frac{\log NA}{J_i} \right)^m (1 + o(1)). \quad (\text{A.1})$$

It suffices to show that this lower bound is, at the same time, asymptotically the upper bound. To this end, we introduce the sequence of auxiliary stopping times

$$\nu_i(\lambda) = \min \{n \geq \lambda : Z_i^\lambda(n) \geq \log NA\}, \quad \lambda = 1, 2, \dots \quad (\text{A.2})$$

It follows from Theorem 4.2 in Dragalin et al. (1999) that for every $m \leq r$

$$\mathbf{E}_{\lambda,i} [\nu_i(\lambda) - \lambda]^m \sim \left(\frac{\log NA}{J_i} \right)^m \quad \text{as } A \rightarrow \infty$$

whenever $(n - \lambda)^{-1} Z_i^\lambda(n)$ converges to J_i r -quickly (condition (17.2.5)). Since $\tau^*(A) \leq \nu_i(\lambda)$ on $\{\tau^*(A) \geq \lambda\}$ (see (A.5)–(A.7) below), we have

$$\begin{aligned} \mathbf{E}_{\lambda,i} [(\tau^* - \lambda)^m | \tau^* \geq \lambda] &\leq \mathbf{E}_{\lambda,i} [(\nu_i(\lambda) - \lambda)^m | \tau^* \geq \lambda] = \mathbf{E}_{\lambda,i} (\nu_i(\lambda) - \lambda)^m \\ &= \left(\frac{\log NA}{J_i} \right)^m (1 + o(1)), \end{aligned}$$

where the equality $\mathbf{E}_{\lambda,i} [(\nu_i - \lambda)^m | \tau^* \geq \lambda] = \mathbf{E}_{\lambda,i} (\nu_i - \lambda)^m$ follows from the fact that $\{\tau^* \geq \lambda\} \in \mathcal{F}_{\lambda-1}$ and $\nu_i(\lambda)$ does not depend on $\mathcal{F}_{\lambda-1}$.

Comparing this upper bound with (A.1) completes the proof. ■

Appendix A2. Proof of Theorem 2

In order to prove (17.2.9), we again use an “upper-lower” bound technique.

We first observe that the stopping time τ^* does not exceed the one-sided stopping time $\nu_i(1) = \nu_A$ defined in (A.2), and hence, in order to

derive an upper bound for $D_i^m(\tau^*)$ it is sufficient to find an asymptotic expansion for $\mathbf{E}_i(\nu_A)^m$. Note that in the i.i.d. case the sequence of statistics $\{Z_i(n), n \geq 1\}$ is a random walk with mean $\mathbf{E}_i Z_i(1) = I_i$, which is positive by condition (17.2.8).

The second important observation is that, again by condition (17.2.8) (finiteness), $\mathbf{E}_i\{Z_i^-(1)\}^m < \infty$ for all $m > 0$, where

$$Z_i^-(1) = -\min\{0, Z_i(1)\}.$$

Indeed,

$$\mathbf{E}_i e^{Z_i^-(1)} = \mathbf{E}_i e^{-Z_i(1)} \mathbb{1}_{\{Z_i(1) < 0\}} + \mathbf{E}_i \mathbb{1}_{\{Z_i(1) \geq 0\}} \leq \mathbf{E}_i e^{-Z_i(1)} + 1 = 2.$$

Therefore, Theorem III.8.1 of Gut (1988) can be applied to show that

$$\mathbf{E}_i(\nu_A)^m = \left(\frac{\log NA}{I_i}\right)^m (1 + o(1)) \quad \text{as } A \rightarrow \infty \text{ for all } m \geq 1.$$

Setting $A = T$, we obtain that for all $m \geq 1$

$$D_i^m(\tau^*) \leq \left(\frac{\log NT}{I_i}\right)^m (1 + o(1)) \quad \text{as } T \rightarrow \infty. \tag{A.3}$$

On the other hand, by the strong law of large numbers

$$\lim_{n \rightarrow \infty} \mathbf{P}_i \left\{ \max_{1 \leq k \leq n} Z_i(k) \geq (1 + \varepsilon) I_i n \right\} = 0 \quad \text{for all } \varepsilon > 0,$$

and a slight modification of the proof of Theorem 1 of Lai (1998) applies, thereby showing that

$$\inf_{\tau \in \Delta(T)} D_i^1(\tau) \geq \frac{\log T}{I_i} (1 + o(1)) \quad \text{as } T \rightarrow \infty.$$

Applying Jensen's inequality, we also deduce that for all $m \geq 1$

$$\inf_{\tau \in \Delta(T)} D_i^m(\tau) \geq \left(\frac{\log T}{I_i}\right)^m (1 + o(1)) \quad \text{as } T \rightarrow \infty,$$

which, along with the upper bound (A.3), proves (17.2.9).

To prove (17.2.10), we introduce the statistic $\Lambda(n) = N^{-1} \sum_{i=1}^N e^{Z_i(n)}$ and the auxiliary stopping time $\tau_A = \min\{n : \Lambda(n) \geq A\}$. By \mathbf{E} , we

denote the expectation with respect to the measure $\mathbf{P} = N^{-1} \sum_{i=1}^N \mathbf{P}_i$. We have

$$\begin{aligned} \mathbf{P}_\infty \{ \tau_A < \infty \} &= \mathbf{E} \left[\frac{1}{\Lambda(\tau_A)} \mathbb{1}_{\{ \tau_A < \infty \}} \right] = \frac{1}{A} \mathbf{E} [e^{-\rho_A} \mathbb{1}_{\{ \tau_A < \infty \}}] \\ &= \frac{1}{AN} \sum_{i=1}^N \mathbf{E}_i [e^{-\rho_A} \mathbb{1}_{\{ \tau_A < \infty \}}], \end{aligned}$$

where $\rho_A = \log \Lambda(\tau_A) - \log A$ is the excess of $\Lambda(n)$ over $\log A$ at the stopping time τ_A on $\{ \tau_A < \infty \}$. Clearly,

$$\tau_A = \min \{ n \geq 1 : Z_i(n) + Y_i(n) \geq \log(AN) \},$$

where

$$Y_i(n) = \log \left(1 + \sum_{j \neq i} e^{Z_j(n) - Z_i(n)} \right).$$

Since $Y_i(n) \rightarrow 0$ as $n \rightarrow \infty$ \mathbf{P}_i -a.s., the $Y_i(n)$, $n \geq 1$, are slowly changing under \mathbf{P}_i (see Siegmund (1985) or Woodroffe (1982) for the definition of slowly changing sequences). Therefore, Theorem 4.1 in Woodroffe (1982) applies to show that $\rho_A \rightarrow \varkappa_i$ as $A \rightarrow \infty$ in \mathbf{P}_i -distribution, where $\varkappa_i = Z_i(\nu_i) - a$ is the overshoot in the one-sided test ν_i . Also, since $Z_i(n)/n \rightarrow I_i$ \mathbf{P}_i -a.s., $\sup_n Z_i(n) = \infty$ with probability 1, which implies that $\mathbf{P}_i \{ \tau_A < \infty \} = 1$. Therefore,

$$A \mathbf{P}_\infty \{ \tau_A < \infty \} \xrightarrow{A \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \gamma_i = \bar{\gamma}_N. \tag{A.4}$$

Finally, generalizing the quite tedious argument used in Pollak (1987), we are able to show that

$$\mathbf{E}_\infty R(\tau^*) = \mathbf{E}_\infty \tau^* \sim \frac{1}{\mathbf{P}_\infty \{ \tau_A < \infty \}} \quad \text{as } A \rightarrow \infty,$$

which, along with (A.4), completes the proof of (17.2.10).

Consider the hypothesis $H_{i,1}$. Observe that the statistic $\log R(n)$ can be represented in the form

$$\log R(n) = Z_i(n) + V_{i,n} - \log N, \tag{A.5}$$

where

$$V_{i,n} = \log \left(1 + \sum_{s=1}^{n-1} \exp \{-Z_i(s)\} + \exp \{-Z_i(n)\} \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{s=1}^n \exp \left\{ \sum_{k=s}^n Z_j^s(n) \right\} \right). \quad (\text{A.6})$$

Therefore, the stopping time $\tau^*(A)$ can be written as

$$\tau^*(A) = \min \{n \geq 1 : Z_i(n) + V_{i,n} \geq \log(AN)\}. \quad (\text{A.7})$$

The sequence $V_{i,n}$ is slowly changing under \mathbf{P}_i and converges in \mathbf{P}_i -distribution to a random variable

$$V_i = \log \left(1 + \sum_{k=1}^{\infty} e^{-Z_i(k)} \right)$$

with the expectation

$$\mathbf{E}_i V_i = C_i = \mathbf{E}_i \left\{ \log \left(1 + \sum_{k=1}^{\infty} e^{-Z_i(k)} \right) \right\}.$$

Thus, the proof of (17.2.11) can be completed by using Woodrooffe's nonlinear renewal theorem (see Woodrooffe (1982), Theorem 4.5). To use this theorem, we have to check the conditions (4.13) and (4.16) in that monograph, which is a straightforward but tedious task. Note that condition (4.14) in Woodrooffe (1982) holds trivially, since $V_{i,n}$ are nonnegative. Applying this theorem gives (17.2.11) ■

ACKNOWLEDGMENT

We wish to thank Dr. John Barnett and Dr. Steven Doss-Hammel of SPAWAR Systems Center, San Diego for useful comments and providing real IR data. We are also grateful to Dr. Boris Rozovsky and Dr. Vlad Repin for valuable discussions and help in the work. Application to IRST is our joint work with Rudolf Blažek (clutter rejection), Anton Petrov and Boris Rozovsky (track-before-detect). Application to network security is the joint work with Boris Rozovsky, Rudolf Blažek, and Hongjoong Kim. Finally, we are grateful to reviewers for valuable comments.

The research of A. Tartakovsky was supported in part by the U.S. ONR grants N00014-99-1-0068 and N00014-95-1-0229 and by the U.S. DARPA grant N66001-00-C-8044. The research of V.V. Veeravalli was supported in part by the U.S. ONR grant N00014-97-1-0823 and by the U.S. NSF CAREER/PECASE grant CCR-0049089.

REFERENCES

- [1] Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall: Englewood Cliffs.
- [2] Blažek, R., Kim, H., Rozovskii, B., and Tartakovsky, A. (2001). A novel approach to detection of “denial-of-service” attacks via adaptive sequential and batch-sequential change-point detection methods. In *Proc. IEEE Systems, Man, and Cybernetics Information Assurance Workshop*, West Point, New York.
- [3] Blažek, R.B., Kim, H., Rozovskii, B., and Tartakovsky, A. (2003). The quickest sequential detection of intrusions in computer networks. *Interface 2003*.
- [4] CERT. (1996). *TCP SYN Flooding and IP Spoofing Attacks*. CERT Advisory CA-96.21.
- [5] Dragalin, V.P. (1995). A multi-channel change point problem. In *Proc. 3rd Umea-Wuzburg Conf. in Statistics*, Umea University, 97–108.
- [6] Dragalin, V.P. (1996). Adaptive procedures for detecting a change in distribution. In *Proc. 4th Wuerzburg-Umea Conf. in Statistics*, Umea University, 87–103.
- [7] Dragalin, V.P., Tartakovsky, A.G., and Veeravalli, V. (1999). Multihypothesis sequential probability ratio tests, I: asymptotic optimality. *IEEE Trans. Inform. Theory*, **45**, 2448–2461.
- [8] Gut, A. (1988). *Stopped Random Walks: Limit Theorems and Applications*. Springer-Verlag: New York.
- [9] Hsu, P.L. and Robbins, H. (1947). Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. U.S.A.*, **33**, 25–31.

- [10] Kent, S. (2000). On the trial of intrusions into information systems. *IEEE Spectrum*, 52–56.
- [11] Kligys, S., Rozovsky, B.L., and Tartakovsky, A.G. (1998). Detection algorithms and track before detect architecture based on nonlinear filtering forIRST. *Technical report CAMS-98.9.1*, Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, U.S.A. <http://www.usc.edu/dept/LAS/CAMS/usr/facmemb/tartakov>
- [12] Lai, T.L. (1976). On r -quick convergence and a conjecture of Strassen, *Ann. Probab.*, **4**, 612–627.
- [13] Lai, T.L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc., Ser. B*, **57**, No. 4, 613–658.
- [14] Lai, T.L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory*, **44**, 2917–2929.
- [15] Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.*, **42**, 1987–1908.
- [16] Nikiforov, I.V. (1995). A generalized change detection problem. *IEEE Trans. Inform. Theory*, **41**, 171–187.
- [17] Petrov, A., Rozovskii, B.L. and Tartakovsky, A.G. Efficient nonlinear filtering methods for detection of dim targets by passive systems. *IEEE Trans. Aerosp. Elec. Sys.* (submitted).
- [18] Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.*, **13**, 206–227.
- [19] Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.*, **15**, 749–779.
- [20] Rozovskii, B. and Petrov, A. (1999). Optimal nonlinear filtering for track-before-detect in IR image sequences. In *SPIE Proceedings: Signal and Data Processing of Small Targets*, **3809**, 152–163.

- [21] Shiryaev, A.N. (1963). On optimum methods in quickest detection problems. *Theor. Probab. Appl.*, **8**, 22–46.
- [22] Shiryaev, A.N. (1978). *Optimal Stopping Rules*. Springer-Verlag: New York.
- [23] Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag: New York.
- [24] Tartakovsky, A.G. (1988). Multi-alternative sequential detection and estimation of signals with random appearance times. In *Statist. Control Problems*, **83**, 216–222.
- [25] Tartakovsky, A.G. (1991). *Sequential Methods in the Theory of Information Systems*. Radio i Svyaz': Moscow (In Russian).
- [26] Tartakovsky, A.G. (1991). Asymptotically optimal multi-alternative sequential detection of a disorder of information systems. In *Proc. IEEE Intern. Symp. Inform. Theory*, Budapest, 359-.
- [27] Tartakovskii, A.G. (1992). Efficiency of the generalized Neyman-Pearson test for detecting changes in a multichannel system. *Probl. Inform. Transmis.*, **28**, 341–350.
- [28] Tartakovsky, A.G. (1994). Asymptotically minimax multialternative sequential rule for disorder detection. In *Statistics and Control of Random Processes: Proc. Steklov Institute of Mathematics*, **202**, Issue 4, 229–236. AMS, Providence, Rhode Island.
- [29] Tartakovsky, A.G. (2003). Extended asymptotic optimality of certain change-point detection procedures. *Ann. Statist.* (submitted).
- [30] Tartakovsky, A.G. and Blažek, R. (2000). Effective adaptive spatial-temporal technique for clutter rejection inIRST. In *SPIE Proceedings: Signal and Data Processing of Small Targets*, **4048**, 85–95.
- [31] Tsitsiklis, J.N. (1993). Extremal properties of likelihood-ratio quantizers. *IEEE Trans. Commun.*, **41**, no. 4, 550–558.

[32] Veeravalli, V.V. (2001). Decentralized quickest change detection. *IEEE Trans. Inform. Theory*, **47**, 1657–1665.

[33] Woodroffe, M. (1982) *Nonlinear Renewal Theory in Sequential Analysis*. SIAM: Philadelphia.

Addresses for Communication:

ALEXANDER TARTAKOVSKY, Center for Applied Mathematical Sciences, University of Southern California, 1042 Downey Way, DRB-155, Los Angeles, CA 90089-1113, U.S.A., Email: tartakov@math.usc.edu
VENUGOPAL VEERAVALLI, ECE & CSL, University of Illinois at Urbana-Champaign, 1308 West Main Street, Urbana, IL 61801, U.S.A., Email: vvv@uiuc.edu