Asymptotic Statistical Properties of Communication-Efficient Quickest Detection Schemes in Sensor Networks

Ruizhi Zhang & Yajun Mei
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, Georgia, USA

Abstract: The quickest change detection problem is studied in a general context of monitoring a large number K of data streams in sensor networks when the "trigger event" may affect different sensors differently. In particular, the occurring event might affect some unknown, but not necessarily all, sensors, and also could have an immediate or delayed impact on those affected sensors. Motivated by censoring sensor networks, we develop scalable communication-efficient schemes based on the sum of those local CUSUM statistics that are "large" under either hard, soft, or order thresholding rules. Moreover, we provide the detection delay analysis of these communication-efficient schemes in the context of monitoring K independent data streams, establish their asymptotic statistical properties under two regimes: one is the classical asymptotic regime when the dimension K is fixed, and the other is the modern asymptotic regime when the dimension K goes to ∞. Our theoretical results illustrate the deep connections between communication efficiency and statistical efficiency.

Keywords: Asymptotic optimality; Change point; Detection delay; False alarm rate; High-dimensional data.

Subject Classifications: 62L15; 60G40.

1. INTRODUCTION

Sensor networks have broad applications including health and environmental monitoring, biomedical signal processing, wireless communication, intrusion detection in computer networks, and surveillance for national security. There are many important dynamic decision problems in sensor networks, as information is accumulated (or updated) over time in the network systems. One of them is the quickest detection of a "trigger" event when sensor networks are deployed to monitor the changing environments over time and space, see Veeravalli (2001).

In this article, we consider a general scenario of quickest detection problems when some unknown, but not necessarily all, sensors might be affected by the "trigger event." A naive approach is to monitor each local sensor individually and to raise a global alarm as soon as any local sensor raises a local alarm. Unfortunately, this specific parallel local monitoring approach does not take

Address correspondence to Yajun Mei, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive NW, Atlanta, GA 30332-0205, USA; E-mail: ymei@isye.gatech.edu

advantage of global information and may lead to large detection delays if several sensors can provide information about the occurring event. Indeed, one allegation often made to the parallel local monitoring approach is that one loses much information at the global level by combining local detection procedures, not raw observations themselves, to make a global decision.

The main purpose of this article is to demonstrate that the problem is not on the parallel local monitoring approach itself, but on how to combine the local detection statistics suitably when the number of affected data streams is moderate. Our proposed methodologies are motivated by the communication efficiency in censoring sensor network, which was introduced by Rago et al. (1996) and later by Appadwedula et al. (2005) and by Tay et al. (2007). Figure 1 illustrates the general setting of a widely used configuration of censoring sensor networks, in which the data streams $X_{k,n}$'s are observed at the remote sensors (typically low-cost battery-powered devices), but the final decision is made at a central location, called the fusion center. The key feature of such a network is that while sensing (i.e., taking observations at the local sensors) is generally cheap and affordable, communication between remote sensors and fusion center is expensive in terms of both energy and limited bandwidth. Thus, to prolong the reliability and lifetime of the network system, practitioners often allow the local sensors to send summary messages $U_{k,n}$'s to the fusion center only when necessary. The question then becomes when and how to send summary messages so that the fusion center can still monitor the network system effectively.

This consideration motivates us to propose communication-efficient schemes that raise a global alarm based on the sum of those local detection statistics (e.g., local CUSUM statistics) that are "large" under either hard-, soft- or order- thresholding. We will then investigate the statistical properties of our proposed communication-efficient schemes under two asymptotic regimes: one is the classical asymptotic regime for fixed dimension K, and the other is the modern asymptotic regime when the dimension K goes to ∞. Our theoretical results illustrate the deep connections between communication efficiency and statistical efficiency.

It is worth pointing out that a well-known view in the standard off-line statistical inference literature is the necessity of shrinkage or thresholding for high-dimensional data in order to improve statistical power or efficiency, see Candès (2006) and the references there. In the sequential change-point detection or quickest detection literature, shrinkage or thresholding has been applied in two different directions for sparse post-change scenarios: one direction is the application on the shrinkage estimation of sparse post-change parameters of local data streams, see Xie and Siegmund (2013); Wang and Mei (2015); Chan (2017), and the other is an indirect approach of filtering out non-changing local data streams through the local summary statistics, which was first proposed in a conference paper by the author in Mei (2011) and were shown to be effective in real-world applications of profile or image monitoring (Liu et al. (2015); Zhang et al. (2018)). This article investigates the asymptotic statistical properties of the indirect approach, and hopefully it will provide a deeper insight and popularize its use in practice to balance the tradeoff between communication efficiency and statistical efficiency.

The remainder of this article is organized as follows. In Section 2 we present a rigorous mathematical formulation of sequential change-point detection problems in the context of globally monitoring multiple data streams and also discuss existing methodologies. In Section 3, we develop our proposed methodologies from the communication-efficient viewpoint and provide guidelines how to choose tuning parameters. Asymptotic statistical properties of our proposed communication-efficient schemes are presented in section 4 and numerical Monte Carlo simulation results are provided in section 5. The detailed technical proofs are postponed in the appendix.

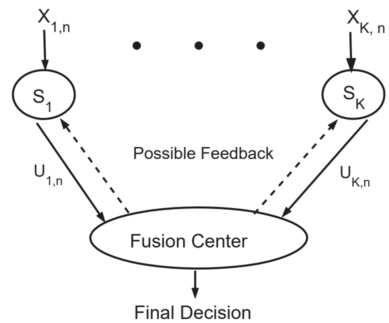

Figure 1: A widely used configuration of censoring sensor networks.

2. PROBLEM FORMULATION AND BACKGROUNDS

Suppose that in a network system as in Figure 1, there are K sensors, and each local sensor S_k observes a local data stream over time, say, $\{X_{k,n}\}_{n=1}^{\infty}$ for $k = 1, \ldots, K$. Initially, the system is "in control" and the distribution of the $X_{k,n}$'s is f_k at the k-th sensor. At some *unknown* time ν, a "trigger" event occurs to the network system, and the density function of the sensor observations $X_{k,n}$'s changes from one density f_k to another density g_k at time $\nu_k = \nu + \delta_k$. Here the term $\delta_k \in [0, \infty]$ denotes the (unknown) delay of the occurring event's impact at the k-th sensor, and $\delta_k = \infty$ implies that the k-th sensor is not affected. The problem is to find an efficient global monitoring scheme, so that the system can detect the occurring event as quickly as possible.

To be more rigorous, we assume that the f_k's and g_k's are completely specified densities with respect to a suitable measure μ_k, see, for example, Tartakovsky and Veeravalli (2004). For each $1 \leq k \leq K$, we assume that the Kullback-Leibler (KL) information number

$$I(g_k, f_k) = \int \log \frac{g_k(x)}{f_k(x)} g_k(x) d\mu_k(x) \tag{2.1}$$

is finite and positive, and

$$\int \Big(\log \frac{g_k(x)}{f_k(x)} \Big)^2 g_k(x) d\mu_k(x) < \infty. \tag{2.2}$$

Denoted by $\mathbf{P}^{(\nu)}_{\delta_1,\delta_2,\ldots,\delta_K}$ and $\mathbf{E}^{(\nu)}_{\delta_1,\delta_2,\ldots,\delta_K}$ the probability measure and expectation of the sensor observations when the event occurs at time ν, and denoted by $\mathbf{P}^{(\infty)}$ and $\mathbf{E}^{(\infty)}$ the same when there are no changes. Note that $\mathbf{P}^{(\nu)}_{\infty,\infty,\ldots,\infty}$ is the same as $\mathbf{P}^{(\infty)}$. A global monitoring scheme can be defined as a stopping time T with respect to the sequence of K-dimensional random vectors $\{(X_{1,n}, \cdots, X_{K,n})\}_{n \geq 1}$, and the interpretation of T is that, when $T = n$, we stop at time n and declare that a change has occurred somewhere at or before time n. As in the classical quickest change detection problems in Lorden (1971), our problem can then be formulated as to find a stopping time T such that the "worse-case" detection delay

$$\overline{\mathbf{E}}_{\delta_1,\delta_2,\ldots,\delta_K}(T) = \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbf{E}^{(\nu)}_{\delta_1,\delta_2,\ldots,\delta_K} \Big((T - \nu + 1)^+ \Big| \mathcal{F}_{\nu-1} \Big) \tag{2.3}$$

is as small as possible for those reasonable combinations of nonnegative δ_k's subject to the global false alarm constraint

$$\mathbf{E}^{(\infty)}(T) \geq \gamma, \tag{2.4}$$

where $\gamma > 0$ is a pre-specified constant.

When $K = 1$ or when monitoring a single local data stream, say, the k-th data stream, such a problem has been well studied in the sequential change-point detection literature, see Page (1954); Shiryaev (1963); Lorden (1971); Pollak (1985, 1987); Moustakides (1986); Basseville and Nikiforov (1993); Lai (1995, 2001); Kulldorff (2001). For a review, see the books such as Basseville and Nikiforov (1993), Poor and Hadjiliadis (2009), Tartakovsky et al. (2014). One efficient local detection procedure is Page's CUSUM procedure: it raises a local alarm at the first time n when the local CUSUM statistic $W_{k,n}$ exceeds some pre-specified threshold, where $W_{k,n}$ can be computed conveniently online via a recursive formula

$$\begin{aligned} W_{k,n} &= \max\Big\{0, \max_{1\le\nu\le n} \sum_{i=\nu}^{n} \log \frac{g_k(X_{k,i})}{f_k(X_{k,i})}\Big\} \\ &= \max\Big(W_{k,n-1} + \log \frac{g_k(X_{k,n})}{f_k(X_{k,n})},\ 0\Big). \end{aligned} \tag{2.5}$$

Below we will develop global monitoring schemes based on the local CUSUM statistics $W_{k,n}$ in (2.5), although the ideas can be easily extended to other local detection statistics (in the logarithm scale of the likelihood) such as Shiryeav-Roberts statistics or scan statistics.

Now let us go back to our global monitoring problem when K is moderately large, and it is known that the generalized likelihood ratio based methods do not have recursive forms and are computationally expensive, see Mei (2010). In order to develop efficient scalable global monitoring schemes, it is natural to combine the local detection procedures together to make a global decision, and there are two intuitive approaches. The first one is the "MAX" scheme that raises an alarm at the global level if the maximum of the local CUSUM statistics is too large, i.e., if one of the local CUSUM procedures raises a local alarm, see Tartakovsky et al. (2006). Mathematically, the "MAX" scheme raises a global alarm at time

$$T_{\max}(c) = \inf\{n \ge 1 : \max_{1\le k\le K} W_{k,n} \ge c\}, \tag{2.6}$$

($= \infty$ if such n does not exist) where $c > 0$ is a pre-specified constant chosen to satisfy the false alarm constraint (2.4). The second approach is the "SUM" scheme, proposed in Mei (2010), in which one raises an alarm if the sum of local CUSUM statistics is too large. Specifically, at time n, each data stream calculates its local CUSUM statistic $W_{k,n}$'s as in (2.5), and then one will raise an alarm at the global level at time

$$T_{\text{sum}}(d) = \inf\{n \ge 1 : \sum_{k=1}^{K} W_{k,n} \ge d\}, \tag{2.7}$$

where the constant $d > 0$ is some suitably chosen constant. Intuitively, the "MAX" scheme $T_{\max}(c)$ in (2.6) works better when one or very few data streams are affected, whereas the "SUM" scheme $T_{\text{sum}}(d)$ in (2.7) works better when many data streams are affected, and numerical simulations in Mei (2010) indeed verified this intuition.

3. COMMUNICATION-EFFICIENT METHODOLOGY

In this section, we propose our global monitoring schemes from the communication efficiency viewpoint in the censoring sensor networks in Figure 1. To have a better illustration, we divide this section to two subsections. In the first subsection, we will present our proposed schemes and provides the motivation of our proposed schemes in the censoring sensor networks. In the second subsection, we will discuss the relation between the tuning parameters in our proposed schemes and the communication costs in the censoring sensor networks and provide guidelines about how to choose the tuning parameters.

3.1. Our Proposed Schemes

From the communication efficiency viewpoint, in the censoring sensor networks in Figure 1, the local sensors need to summarize the information and only send "significant" information to the fusion center to prolong the reliability and lifetime of the network. This inspires us to propose to transmit only those local CUSUM statistics $W_{k,n}$'s that are larger than their respective local thresholds.

Specifically, at time n, each local sensor calculates its local CUSUM statistic $W_{k,n}$ recursively as in (2.5), and then sends the following sensor message $U_{k,n}$ to the fusion center:

$$U_{k,n} = \begin{cases} W_{k,n}, & \text{if } W_{k,n} \geq b_k \\ \text{NULL}, & \text{if } W_{k,n} < b_k \end{cases}, \tag{3.1}$$

where $b_k \geq 0$ is the local censoring (hard threshold) parameter at the k-th sensor. Here the message "NULL" is a special sensor symbol to indicate the local CUSUM statistic is not large. In practice, "NULL" could be represented by the situation when the sensor does not send any messages to the fusion center, e.g., the sensor is silent.

After receiving the local sensor messages $U_{k,n}$'s in (3.1), the fusion center then combines them together suitably to make a global decision. There are several reasonable approaches to do so, and the first two schemes are based on the summation of all sensor messages $U_{k,n}$'s, depending on how to interpret the "NULL" values. The first approach is to treat the "NULL" values as lower limit 0, and to raise a global alarm at the fusion center at time

$$\begin{aligned} N_{hard}(a) &= \inf\Big\{n \geq 1 : \sum_{k=1}^{K} U_{k,n} \geq a\Big\} \\ &= \inf\Big\{n \geq 1 : \sum_{k=1}^{K} W_{k,n}\mathbf{1}\{W_{k,n} \geq b_k\} \geq a\Big\}. \end{aligned} \tag{3.2}$$

Below this scheme will be referred as the hard-thresholding scheme, since it involve the hard-thresholding transformation $h(w) = w\mathbf{1}\{w \geq b\}$ of the local CUSUM statistics $W_{k,n}$.

The second approach is to treat the "NULL" values as the upper limit b_k's, in which the fusion center will compute the global monitoring statistic

$$G_n = \sum_{k=1}^{K} U_{k,n} = \sum_{k=1}^{K} \max\{W_{k,n}, b_k\} = \sum_{k=1}^{K} \max\{W_{k,n} - b_k, 0\} + \sum_{k=1}^{K} b_k.$$

This is closely related to the soft-thresholding transformation $h(w) = \max(w - b, 0)$ of the local CUSUM statistic $W_{k,n}$, and we can define the soft-thresholding scheme that raises an alarm at time

$$N_{soft}(a) = \inf\Big\{n \geq 1 : \sum_{k=1}^{K} \max\{W_{k,n} - b_k, 0\} \geq a\Big\}. \tag{3.3}$$

Here we keep the threshold of $N_{soft}(a)$ as a instead of $a - \sum_{k=1}^{K} b_k$, so that both $N_{hard}(a)$ in (3.2) and $N_{soft}(a)$ in (3.3) can be written in a common SUM-shrinkage family of schemes

$$N_G(a) = \inf\{n \geq 1 : \sum_{k=1}^{K} h_k(W_{k,n}) \geq a\}, \tag{3.4}$$

also see Liu et al. (2018).

The third approach occurs when the fusion center has a prior knowledge that (at most) r out of K data streams will be affected by the occurring event. Such a prior knowledge may be defined by the network fault-tolerant design to avoid risking failure. In this case, it is reasonable for the fusion center to order all sensor messages $U_{k,n}$'s as $U_{(1),n} \geq \ldots \geq U_{(K),n}$, and raise an alarm if the sum of the r largest $U_{k,n}$'s is too large. This yields a global monitoring scheme that is based on the order-thresholding transformation of $U_{k,n}$'s:

$$N_{comb,r}(a) = \inf\Big\{n \geq 1 : \sum_{k=1}^{r} U_{(k),n} \geq a\Big\}, \tag{3.5}$$

where one might treats the "NULL" values as lower limit 0, upper limit b_k or any other reasonable values. In this article, $U_{k,n}$ in the combined scheme $N_{comb,r}(a)$ is chosen as the hard-shrinkage of the local CUSUM statistics, i.e., $W_{k,n}\mathbf{1}\{W_{k,n} \geq b_k\}$.

From the statistical viewpoint, a special case of $N_{comb,r}(a)$ in (3.5) is when the order-thresholding transformation is applied directly to the local detection statistics $W_{k,n}$'s in (2.5) themselves. Specifically, we order the K local CUSUM statistics $W_{1,n}, \ldots, W_{K,n}$ from largest to smallest: $W_{(1),n} \geq W_{(2),n} \geq \ldots \geq W_{(K),n}$. Then the order-thresholding scheme can be defined by the stopping time

$$N_{order,r}(a) = \inf\Big\{n \geq 1 : \sum_{k=1}^{r} W_{(k),n} \geq a\Big\}. \tag{3.6}$$

Clearly, $N_{order,r}(a)$ is a special case of $N_{comb,r}(a)$ if the local censoring parameter $b_k \equiv 0$, since the local CUSUM statistics $W_{k,n}$'s are non-negative.

Note that each family of schemes, $N_{hard}(a)$ in (3.2), $N_{soft}(a)$ in (3.3), $N_{order,r}(a)$ in (3.6), and $N_{comb,r}(a)$ in (3.5), can be thought of as a large family that includes both "MAX" and "SUM" schemes. For instance, the "SUM" scheme $T_{\text{sum}}(d)$ in (2.7) correspond to the hard thresholding scheme $N_{hard}(a)$ with $b_k \equiv a$ and $a = d$, or the order-thresholding scheme $N_{order,r}(a)$ in (3.6) with $r = 1$. Similarly, if all threshold parameter $b_k = 0$, then the hard thresholding scheme $N_{hard}(a)$ in (3.2), the soft-thresholding schemes $N_{soft}(a)$, and $N_{comb,r}(a)$ in (3.5) with $r = K$ will become the "SUM" scheme $T_{\text{sum}}(d)$ in (2.7).

It is useful to mention that our proposed schemes, $N_{hard}(a)$ in (3.2), $N_{soft}(a)$ in (3.3), $N_{order,r}(a)$ in (3.6), and $N_{comb,r}(a)$ in (3.5), take advantage of the same high-level insights: little information seems to be lost at the fusion center if we do not observe those local data streams with small values of $W_{k,n}$'s since they make limited contributions to detect the true changes. These ideas and similar techniques have been applied in other contexts. Banerjee and Veeravalli (2015) essentially use the hard-thresholding transformation in (3.2) tackle the quickest detection problem when one purposely miss the observations to reduce costs. Wang et al. (2018) borrowed the soft-threshold schemes in (3.3) for profile monitoring when a change only affects some but not all principle components in the principal component analysis. Liu et al. (2015) applied the order-thresholding transformation in (3.6) for efficient adaptive sampling policy when one only has ability to observe r out of K data streams at each time step. This may occur in manufacturing process control when there are K possible stages in the process but there are only r expensive sensors available to monitor the process. In such a problem, the order-thresholding scheme allows us to adaptively observe those r data streams with the largest $W_{k,n}$'s values at each time step. Zhang et al. (2018) also used the order-thresholding transformation in (3.6) for monitoring nonlinear profiles when small shifts may occurred on some unknown regions of the profile data. In addition, along the idea of order statistics, Banerjee and Fellouris (2016) proposed the stopping time $\hat{N}_r(a) = \inf\{n : W_{(r),n} \geq a\}$. This is asymptotically equivalent to our proposed order-thresholding scheme $N_{order,r}(a)$ in (3.6) when the prior knowledge of exactly r affected data streams is true. However, our proposed order-thresholding scheme $N_{order,r}(a)$ in (3.6) is more robust when the prior knowledge is inaccurate, particularly when the true affected number of data streams $r_{true} < r$.

3.2. Choice of Thresholding Parameters

So far we simply follow our intuition without discussing how to choose the local threshold parameters b_k's. Intuitively we should choose identical local threshold parameters b_k's when the local sensors are homogeneous, but choose sensor-specified local threshold parameters b_k's when the sensors are nonhomogeneous. The homogeneous case was discussed in our previous research in Liu et al. (2018), and here we focus on the possible nonhomogeneous case.

Under the assumption of the finiteness of local KL information numbers $I(g_k, f_k)$ in (2.1), we propose to choose the local threshold parameter b_k's as

$$b_k = \rho_k b \tag{3.7}$$

for $k = 1, \dots, K$, where

$$\rho_k = \frac{I(g_k, f_k)}{\sum_{k=1}^{K} I(g_k, f_k)} \tag{3.8}$$

and $b \geq 0$ is the common global-level thresholding parameter that will be discussed in a little bit. The rigorous statistical justification of (3.7)-(3.8) will be postponed to the next section, and it is useful to think at the high-level that ρ_k can be thought of as the weight of the k-th data stream in the overall final decision, and those local sensors with larger KL information numbers or larger signal-to-noise ratios will play more important roles in the final decision. Meanwhile, note that when the sensors are the homogeneous, we have $\rho_k \equiv 1/K$ and thus local threshold parameters $b_k \equiv b/K$ are the same. Hence, our proposed choices of thresholding parameters in (3.7)-(3.8) match our intuition in the homogeneous case.

The choice of global-level thresholding parameter b is nontrivial, and may need to consider some non-statistical constraints. As an illustration, in certain applications of censoring sensor networks, the censoring parameter b may be chosen to satisfy the constraints on the average fraction of transmitting sensors when no events occur. For our proposed scheme $N_{hard}(a, b)$, when no event occurs, the average fraction of transmitting sensors at any time step n is

$$\frac{1}{K}\sum_{k=1}^{K}\mathbf{P}^{(\infty)}(U_{k,n}\neq \text{NULL}) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{P}^{(\infty)}(W_{k,n}\geq \rho_k b) \leq \frac{1}{K}\sum_{k=1}^{K}\exp(-\rho_k b),$$

where the last inequality follows from the well-known properties of the local CUSUM statistics, see, Appendix 2 on Page 245 of Siegmund (1985). In particular, if all K sensors are homogeneous in the sense that the $I(g_k, f_k)$'s are the same for all k, then $\rho_k = 1/K$, and the average fraction of transmitting sensors at any time step is $\exp(-b/K)$ when no event occurs. Hence for our proposed scheme $N_{hard}(a, b)$, a choice of

$$b = K\log\eta^{-1},$$

or equivalently, the local hard threshold $b_k = \rho_k b = b/K = \log\eta^{-1}$, will guarantee that on average, at most $100\eta\%$ of K homogeneous sensors will transmit messages at any given time when no event occurs. It is interesting to note that the local threshold $b_k = \log\eta^{-1}$ at each local sensor is a constant that does not depend on K.

The choice of b becomes more complicated for the combined thresholding schemes $N_{comb,r}(a, b)$ if the thresholding parameter r has been given beforehand. We do not have an explicit answer, and a general rule of thumb is that the censoring parameter b in (3.5) shall not be too large, as one generally should keep at least r non-zero $U_{k,n}$'s when r data streams are affected by the event.

The choice of thresholding parameter r is straightforward and depends on whether one has any prior knowledge about the maximum number of affected data streams. If such a knowledge exists and it is believed that at most r_0 data streams will be affected by the occuring event, then one should use this r_0 as the value of thresholding parameter r. Otherwise one may want to be conservative to choose $r = K$, e.g., consider the "SUM" scheme or the hard-thresholding scheme $N_{hard}(a, b)$ in (3.2).

4. STATISTICAL EFFICIENCY

In this section, we investigate the statistical efficiency of our proposed communication-efficient schemes, $N_{hard}(a)$ in (3.2), $N_{soft}(a)$ in (3.3), $N_{order,r}(a)$ in (3.6), and $N_{comb,r}(a)$ in (3.5). Here we assume that the local thresholds ρ_k are given in (3.7)-(3.8), and rewrite our proposed schemes as $N_{hard}(a, b)$, $N_{soft}(a, b)$, $N_{order,r}(a, b)$, $N_{comb,r}(a, b)$ so as to emphasize the role of the common threshold b in (3.7). Our statistical efficiency analysis allows us to provide a rationale justification of the choice of ρ_k in (3.8), or b_k in (3.7)-(3.8), although we should emphasize that these choices are a sufficient but not necessarily necessary condition in order for our proposed schemes in (3.2)-(3.6) to enjoy good properties.

For easy understanding our theoretical results, we divide this section into three subsections. In the first subsection, we provide the asymptotic upper bound of detection delay of our proposed schemes under the settings when the number of affected data streams are fixed. In the second subsection, we derive the upper bound of detection delay of our proposed scheme when the false alarm constraint (2.4) γ goes to ∞ under the classical asymptotic regime when the number of data streams K is fixed. The delay analysis on the high-dimension regime when K goes to ∞ will be presented in the last subsection.

4.1. Detection Delay Analysis

In this subsection, we consider a general setting when the change is not necessarily instantaneous. We assume that when the occurring event occurs at time ν, the k-th data stream is affected at time $\nu_k = \nu + \delta_k$, where the term $\delta_k \in [0, \infty]$ denotes the delay of the occurring event's impact on the k-th data stream. In particular, $\delta_k = \infty$ implies that the k-th data stream is not affected. In other words, the density function of the sensor observations $X_{k,n}$'s of the k-th data stream changes from f_k to g_k at time $\nu_k = \nu + \delta_k$. Most research in the literature assumes that the delay effect δ_k only takes two possible values, 0 or ∞. Here we relax such an assumption a little bit, and assume that the delay effects δ_k's satisfy the following post-change hypothesis set Δ :

$$\Delta = \Big\{(\delta_1, \ldots, \delta_K) : \text{the } \delta_k\text{'s either} = \infty \text{ or satisfy } 0 \le \delta_k << \log \gamma \text{ and } \min_{1\le k\le K} \delta_k = 0\Big\}. \quad (4.1)$$

where γ is the false alarm constraint in (2.4), and $x(t) << y(t)$ implies that $x(t)/y(t) \to 0$ as $t \to \infty$. Note that the assumption of $\min_{1\le k\le K} \delta_k = 0$ is trivial, since otherwise the system is actually affected by the occurring event at the "new" change-point $\nu' = \nu + \min_{1\le k\le K} \delta_k$. The assumption of $\delta_k << \log \gamma$ is a technical assumption to ensure that one is able to utilize all affected data streams to raise a global alarm subject to the false alarm constraint γ in (2.4). In other words, we only consider the scenario when the differences on the finite delay affects δ_k's are not too large as compared to the typical order $(\log \gamma)$ of detection delays. A sufficient condition to satisfy this assumption is when all finite δ_k's are uniformly bounded by some constants that do not depend on the false alarm constraint γ in (2.4).

In the detection delay analysis, the following constant plays a crucial role:

$$J(\delta_1, \ldots, \delta_K) = \sum_{k=1}^{K} I(g_k, f_k) I\{\delta_k < \infty\}, \quad (4.2)$$

and $I(g_k, f_k)$ is the KL information number defined in (2.1), and $I\{A\}$ is the indicator function of set A. Essentially, the constant $J(\delta_1, \ldots, \delta_K)$ in (4.2) states that only those affected data streams can make contributions in quickest detection.

The following theorem establishes the detection delay properties of our proposed schemes, $N_{hard}(a, b)$ in (3.2), $N_{soft}(a, b)$ in (3.3), $N_{order,r}(a, b)$ in (3.6), and $N_{comb,r}(a, b)$ in (3.5), as the global threshold a goes to ∞. The proof of this theorem is presented in detail in the appendix.

Theorem 4.1. *Suppose $a \to \infty$.*

(i) *For any combination $(\delta_1, \ldots, \delta_K) \in \Delta$ defined in (4.1), as $b \to \infty$*

$$\begin{aligned} \overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a, b)) \quad &\le \quad \max\left\{\frac{a}{J(\delta_1, \ldots, \delta_K)}, \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\right\} \\ &\quad + O(\sqrt{b}) + O\Big(\max_{\delta_k:\delta_k<\infty}(\delta_k)\Big), \end{aligned} \quad (4.3)$$

where $J(\delta_1, \ldots, \delta_K)$ is defined in (4.2).

(ii) *For all $b \geq 0$, the soft-thresholding scheme $N_{soft}(a, b)$ in (3.3) satisfies*

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{soft}(a,b)) \leq \frac{a}{J(\delta_1,\ldots,\delta_K)} + \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + O(\sqrt{b}) + O\Big(\max_{\delta_k:\delta_k<\infty}(\delta_k)\Big), \tag{4.4}$$

(iii) *For any integer $1 \leq r \leq K$, the order-r thresholding scheme $N_{order,r}(a)$ in (3.6) and the combined thresholding scheme $N_{comb,r}(a,b)$ in (3.5) satisfy (4.3) whenever $\sum_{k=1}^{K} I\{\delta_k < \infty\} \leq r$, i.e., when the occurring event affects at most r sensors.*

4.2. Classical Asymptotic Regime with Fixed Dimension K

In this subsection, we present the asymptotic optimality properties of our proposed schemes, $N_{hard}(a,b)$, $N_{soft}(a,b)$, $N_{order,r}(a)$, and $N_{comb,r}(a,b)$, under the classical asymptotic regime in which the number of data streams K is fix and the false alarm constraint γ goes to ∞.

The following lemma derives the information bound on the detection delays of any globally monitoring schemes when Δ is defined in (4.1), as the false alarm constraint γ in (2.4) goes to ∞.

Lemma 4.1. *Assume a scheme $T(\gamma)$ satisfies the false alarm constraint (2.4). Then for any given post-change hypothesis $(\delta_1,\ldots,\delta_K) \in \Delta$, as γ goes to ∞,*

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(T(\gamma)) \geq (1 + o(1))\frac{\log\gamma}{J(\delta_1,\ldots,\delta_K)}, \tag{4.5}$$

where $J(\delta_1,\ldots,\delta_K)$ is defined in (4.2).

When the local censoring parameters b_k's are defined in (3.7)-(3.8) with the common parameter b, the asymptotic optimality properties of our proposed schemes under the classical asymptotic regime can be summarized as follow.

Theorem 4.2. *For a given K and for any $b \geq 0$, with the choice of*

$$a = a_\gamma = \log\gamma + (K - 1 + o(1))\log\log\gamma, \tag{4.6}$$

the hard-thresholding scheme $N_{hard}(a_\gamma, b)$ satisfies the false alarm constraint (2.4). Moreover, if $a - b$ goes to ∞ as γ goes to ∞, then for all $b \geq 0$,

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a,b)) \leq \frac{\log\gamma + (K-1+o(1))\log\log\gamma}{J(\delta_1,\ldots,\delta_K)} + O(\sqrt{b}) + O(1) \tag{4.7}$$

for all possible post-change hypothesis $(\delta_1,\ldots,\delta_K) \in \Delta$ in (4.1). Therefore, for any given $b = o((\log\log\gamma)^2)$, the hard-thresholding schemes $N_{hard}(a,b)$ in (3.2) asymptotically minimize $\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a,b))$ (up to the second-order) for each and every post-change hypothesis $(\delta_1,\ldots,\delta_K) \in \Delta$ subject to the false alarm constraint (2.4), as γ in (2.4) goes to ∞. The conclusion also holds if $N_{hard}(a,b)$ is replaced by the soft-thresholding scheme $N_{soft}(a,b)$ in (3.3), the order-thresholding scheme $N_{order,r}$ in (3.6) or the combined thresholding scheme $N_{comb,r}(a,b)$ in (3.5) when the occurring event affects at most r data streams, i.e., when $(\delta_1,\ldots,\delta_K) \in \Delta$ satisfies $\sum_{k=1}^{K} I\{\delta_k < \infty\} \leq r$.

Theorem 4.2 validated our choices of the local censoring parameters b_k's in (3.7) and the weights ρ_k's in (3.8) in the general nonhomogeneous scenario, as the corresponding schemes are asymptotically optimal when the KL information numbers $I(g_k, f_k)$ in (2.1) might be different for different k. Moreover, by Theorem 4.2, when $b = o((\log\log\gamma)^2)$, the upper bound of the detection delay in the right hand side of (4.7) is asymptotically first-order equivalent to those with $b = 0$. This indicates that we can choose the local threshold $b = o((\log\log\gamma)^2)$ to achieve both communication efficiency and statistical efficiency simultaneously.

4.3. Modern Asymptotic Regime When the Dimension $K \to \infty$

In this subsection, we present the asymptotic properties of our proposed schemes, $N_{hard}(a, b)$, $N_{soft}(a, b)$, $N_{order,r}(a)$, and $N_{comb,r}(a, b)$, under the modern asymptotic regime in which both the dimension K and the false alarm constraint γ in (2.4) go to ∞ in a suitable rate. In order to be tractable, we consider the homogenous case when $(f_k, g_k) = (f, g)$ for all k, and the local censoring parameters b_k's defined in (3.7)-(3.8) will become $b_k = b/K$ with the common parameter b. In this subsection, denote by $I = I(g, f)$ the KL information number defined in (2.1).

Here we consider the sparse post-change scenario when the number of affected data streams m is fixed, and focus on the impact of the dimension K on the performance of our proposed schemes. Two different scenarios will be investigated: $K = o(\log\gamma)$ and $K >> \log\gamma$. When K and $\log\gamma$ have the same order, research becomes more challenging and is out of the scope of this article. Note that Chan (2017) considers the not-so-sparse and not-so-dense post-change scenario when the number of affected data streams m goes to ∞ by assuming that $\log(m)$, $\log(K)$, and $\log\log\gamma$ have the same order. Here our asymptotic setting is different, and we consider the case of fixed m when K and $\log\gamma$ go to ∞.

First, when both the dimension K and the false alarm constraint γ in (2.4) go to ∞, the choice of a in (4.6) for fixed K might no longer work, and thus it is crucial to find the threshold a to satisfy the false alarm constraint γ in (2.4) in the modern asymptotic setting when $K \to \infty$. The following theorem characterizes a general non-asymptotic result on the conservative choice of the threshold a.

Theorem 4.3. *For any given* b *and* K, *a choice of*

$$a = \left(\sqrt{\log(4\gamma) + K - Ke^{-b/K}} + \sqrt{K}\right)^2 \tag{4.8}$$

will guarantee the hard-shrinkage scheme $N_{hard}(a, b)$, *the soft-thresholding scheme* $N_{soft}(a, b)$, *the order-thresholding scheme* $N_{order,r}(a, b)$ *or the combined thresholding scheme* $N_{comb,r}(a, b)$ *satisfy the false alarm constraint (2.4).*

It is clear from Theorem 4.3 that the asymptotic property of the conservative threshold a in (4.8) depends on the relation between K and $\log\gamma$. The following corollary summarizes the asymptotic detection delays of our proposed schemes, and it shows that the classical asymptotic detection delay bounds for fixed K still hold when $K = o(\log\gamma)$, but we will have new asymptotic delay bounds when $K >> \log\gamma$.

Corollary 4.1. *Assume the number* m *of affected data streams is fixed, and assume* K *and* $\log\gamma$ *go to* ∞,

(i) *if* $K = o(\log \gamma)$, *for any* $b \geq 0$, *with the choice of*

$$a = a_\gamma = \log(4\gamma) + o(\log \gamma) \tag{4.9}$$

the hard-thresholding scheme $N_{hard}(a, b)$ *in (3.2) satisfies the false alarm constraint in (2.4) and has the detection delay*

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a, b)) \leq (1 + o(1))\frac{\log \gamma}{mI} + O(1), \tag{4.10}$$

for all possible post-change hypothesis $(\delta_1, \ldots, \delta_K) \in \Delta$ *in (4.1).*

(ii) *If* $K >> \log \gamma$ *and* $b \geq 0$, *with the choice of*

$$a = (1 + o(1))K \tag{4.11}$$

the hard-thresholding scheme $N_{hard}(a, b)$ *in (3.2) satisfies the false alarm constraint in (2.4). Moreover, if the local censoring parameters* b_k*'s are not too large, i.e.,* $b_k = o(K)$, *or equivalently, the global censoring parameter* $b = o(K^2)$, *we have*

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a, b)) \leq (1 + o(1))\frac{K}{mI} + O(1), \tag{4.12}$$

for all possible post-change hypothesis $(\delta_1, \ldots, \delta_K) \in \Delta$ *in (4.1).*

(iii) *The conclusions of (i) and (ii) also hold if* $N_{hard}(a, b)$ *is replaced by the soft-thresholding scheme* $N_{soft}(a, b)$ *in (3.3), the order-thresholding scheme* $N_{order,r}$ *in (3.6) or the combined thresholding scheme* $N_{comb,r}(a, b)$ *in (3.5) when the occurring event affects at most* r *data streams, i.e., when* $(\delta_1, \ldots, \delta_K) \in \Delta$ *satisfies* $\sum_{k=1}^{K} I\{\delta_k < \infty\} \leq r$.

5. NUMERICAL SIMULATIONS

In this subsection we report our numerical simulation results to illustrate the usefulness of the proposed schemes in (3.2)-(3.6). Suppose that there are $K = 100$ independent and identical sensors in a system, and the observations at each sensor are iid with mean 0 and variance 1 before the change and with mean 1 and variance 1 after the change if affected. In our simulation study, we simply assume that the change is instantaneous if a sensor is affected, but we do not know which subset of sensors will be affected.

For the purpose of comparison, we conduct numerical simulations for six families of global monitoring schemes:

- the "MAX" scheme $T_{\max}(a)$ in (2.6),
- the "SUM" scheme $T_{\text{sum}}(a)$ in (2.7),
- the order thresholding scheme $N_{order,r}(a)$ in (3.6) with $r = 10$,
- the hard thresholding scheme $N_{hard}(a)$ in (3.2),

- the soft thresholding scheme $N_{soft}(a)$ in (3.3),
- the combined thresholding schemes $N_{comb,r}(a)$ in (3.5) with $r = 10$.

The first three schemes require all local sensors to send all local CUSUM statistics $W_{k,n}$'s values to the fusion center at each and every time step, and corresponds to the case when the local censoring parameter $b_k \equiv 0$ for all $k = 1, \cdots, K$. For order-thresholding in the families of $N_{order,r}(a)$ and $N_{comb,r}(a)$, we choose $r = 10$ to better understand the scenario when 10 out of 100 sensors are affected by the occurring event. For each of the last three schemes in the list, i.e., our three proposed schemes (3.2)-(3.5), we further consider three different values of the local censoring parameters b_k's:

(i) $b_k \equiv 1/2 \approx -\log(0.607)$ for all k,

(ii) $b_k \equiv -\log(0.1) = 2.3026$ for all k,

(iii) $b_k \equiv -\log(0.01) = 4.6052$ for all k.

The choices of these values will guarantee that when no event occurs, on average at most $\eta = 60.7\%, 10\%$, and 1% of $K = 100$ homogeneous sensors will transmit messages at any given time, respectively. Therefore, there are a total of $3 + 3 * 3 = 12$ specific schemes in our numerical simulation study.

For each of these 12 specific schemes $T(a)$, we first find the appropriate values of the global threshold a to satisfy the false alarm constraint $\mathbf{E}^{(\infty)}(T(a)) \approx \gamma = 5000$ (within the range of sampling error). Next, using the obtained global threshold value a, we simulate the detection delay when the change-point occurs at time $\nu = 1$ under several different post-change scenarios, i.e., different number of affected sensors. All Monte Carlo simulations are based on $m = 2500$ repetitions.

Table 1 summarizes our simulated detection delays of these 12 schemes under 8 different post-change hypothesis, depending on the number of affected sensors. From Table 1, among these 12 specific schemes, when a small number $(1 \sim 3)$ of 100 homogeneous sensors are affected by the event, the "MAX" scheme $T_{\max}(a)$ is the best (in the sense of smallest detection delay), the "SUM" scheme $T_{\text{sum}}(a)$ is the worst, and all other schemes are in-between. Similarly, when a large number (20 or more) of 100 homogeneous sensors are affected, the order is reserved: $T_{\text{sum}}(a)$ is the best, $T_{\max}(a)$ is the worst, and all other schemes are in-between. However, when $5 \sim 10$ sensors are affected, the schemes with order-thresholding $r = 10$ yield the smallest detection delays, since they are designed to detect the scenario when 10 sensors are affected by the event. In addition, it is clear from Table 1 that for each given scheme, the fewer affected sensors we have, the larger detection delay it will have. All these results are consistent with our intuition.

It is worth emphasizing that for the families of the hard- and soft- thresholding schemes, $N_{hard}(a)$ in (3.2) and $N_{soft}(a)$ in (3.3), a larger censoring value of b_k actually leads to a smaller detection delay when only a few sensors are affected. This suggests that a larger censoring value b_k may actually be necessary for efficient detection when the affected sensors are sparse.

A surprising and possibly counter-intuitive result in Table 1 is the effect of not so large values of censoring parameters b_k's in finite sample simulations. For instance, the performances of the "SUM" scheme $T_{\text{sum}}(a)$ and the hard thresholding scheme $N_{hard}(a, b_k = 0.50)$ are similar in view of sampling errors. Likewise, the top-r thresholding scheme $N_{order,r=10}(a)$ and the combined

Table 1: A comparison of the detection delays of six families of schemes with $\gamma = 5000$. The smallest and largest standard errors of these 12 schemes are also reported under each post-change hypothesis based on 2500 repetitions in Monte Carlo simulations.

	# sensors affected								
	1	3	5	8	10	20	30	50	100
Smallest standard error	0.18	0.07	0.05	0.03	0.03	0.02	0.01	0.01	0.00
Largest standard error	0.35	0.12	0.07	0.06	0.05	0.04	0.03	0.03	0.03
Schemes with $b_k \equiv 0$									
$T_{\max}(a = 11.27)$	23.3	16.3	14.4	13.0	12.4	10.9	10.2	9.5	8.7
$T_{\text{sum}}(a = 88.66)$	52.1	21.8	14.7	10.3	8.7	5.2	3.9	2.9	2.0
$N_{order,r=10}(a = 44.11)$	34.1	15.5	11.2	8.5	7.5	5.5	4.8	4.1	3.4
Schemes $N_{hard}(a)$ in (3.2) with different positive b_k's									
$N_{hard}(a = 85.60, b_k = 0.50)$	52.9	21.9	14.9	10.3	8.7	5.2	4.0	2.9	2.0
$N_{hard}(a = 52.21, b_k = 2.3026)$	50.6	20.7	13.8	9.6	8.2	5.2	4.2	3.2	2.4
$N_{hard}(a = 26.31, b_k = 4.6052)$	39.8	16.0	11.5	8.8	7.9	5.9	5.2	4.4	3.8
Schemes $N_{soft}(a)$ in (3.3) with different positive b_k's									
$N_{soft}(a = 63.92, b_k = 0.50)$	48.2	20.2	13.7	9.7	8.2	5.1	4.0	3.0	2.0
$N_{soft}(a = 21.56, b_k = 2.3026)$	33.9	15.4	11.2	8.5	7.5	5.3	4.5	3.7	3.0
$N_{soft}(a = 8.29, b_k = 4.6052)$	25.2	13.8	11.1	9.2	8.4	6.7	5.9	5.2	4.4
Schemes $N_{comb,r}(a)$ in (3.5) with $r = 10$ and different positive b_k's									
$N_{comb,r}(a = 44.11, b_k = 0.50)$	34.1	15.5	11.2	8.5	7.5	5.5	4.8	4.1	3.4
$N_{comb,r}(a = 43.88, b_k = 2.3026)$	38.5	16.8	11.7	8.6	7.5	5.5	4.7	4.0	3.3
$N_{comb,r}(a = 26.31, b_k = 4.6052)$	39.8	16.0	11.5	8.8	7.9	5.9	5.2	4.4	3.8

thresholding scheme $N_{comb,r=10}(a, b_k = 0.50)$ also have identical performances. The interpretation in the censoring sensor networks context is as follows: using our proposed communication policy in (3.1), we only need $\exp(-b_k) = \exp(-0.5) = 60.7\%$ of 100 sensors to transmit information to the fusion center at any given time when no event occurs, but we can still be as effective as the full transmission scenario when all sensors transmit information at all time steps. In other words, much communication costs can be saved by our proposed schemes $N_{hard}(a)$ or $N_{comb,r}(a)$ with not so large values of b_k's.

It is also interesting to see the effect of the order-thresholding parameter r in finite sample simulations when the hard-thresholding parameters b_k's are large. From Table 1, when the false alarm constraint γ in (2.4) is only moderately large, e.g., $\gamma = 5000$, the performances of $N_{hard}(a, b_k)$ and $N_{comb,r=10}(a, b_k)$ are identical when $b_k = 4.6052$ — they not only have the same global threshold a, but also have the same detection delays. Intuitively, the stopping time $N_{comb,r}(a, b_k)$ is decreasing as a function of r, and thus we have $N_{hard}(a, b_k) = N_{comb,r=K}(a, b_k) \leq N_{comb,r=10}(a, b_k)$ when $b_k = 4.6052$. So one may wonder why our numerical simulations lead to identical results? One explanation is that with such a choice of $b_k = 4.6052$, when no event occurs, on average there is at most 1 non-zero sensor message received in the fusion center at any given time, and thus there is little difference whether one uses the sum of the largest $r = 10$ sensor messages or uses the sum of all $K = 100$ sensor messages. Hence similar performances are observed in finite-sample simulations.

APPENDIX: TECHNICAL PROOFS

Below we present the detailed proofs to Theorems 4.1, 4.2, 4.3 as well as Lemma 4.1 and Corollary 4.1.

Proof of Theorem 4.1. Let us first focus part (i) on the properties of the hard-thresholding scheme $N_{hard}(a,b)$ in (3.2) with $b \geq 0$ being the common constant for b_k's in (3.7)-(3.8).

To prove relation (4.3), it is clear that the worst-case detection delay of $N_{hard}(a,b)$ occurs at the change-point $\nu = 1$, and thus it suffices to show that $\mathbf{E}^{(\nu=1)}_{\delta_1,\ldots,\delta_K}(N_{hard}(a,b))$ satisfies (4.3). Without loss of generality, we assume that only the first m data steams are affected and no other data streams are affected. To simplify our notation below, denote $\delta_{\max} = \max_{1 \leq i \leq m} \delta_i$. It suffices to show that

$$\mathbf{E}^{(\nu=1)}_{\delta_1,\ldots,\delta_K}(N_{hard}(a,b)) \leq \max\left\{\frac{a}{\sum_{k=1}^m I(g_k, f_k)}, \frac{b}{\sum_{k=1}^K I(g_k, f_k)}\right\} + O(\sqrt{b}) + O(1) + \delta_{\max}, \quad (5.1)$$

for any $b \geq 0$.

The essential idea in the proof of (5.1) is to compare $N_{hard}(a,b)$ with new stopping times that are only based on those affected m data streams. Define a stopping time that is in the form of the one-sided sequential probability ratio test (SPRT):

$$\tau(a,b) = \text{first } n \text{ such that } \sum_{i=1}^n \sum_{k=1}^m \log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq a \ \text{ and}$$
$$\sum_{i=1}^n \log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq \rho_k b \ \text{ for all } 1 \leq k \leq m, \quad (5.2)$$

where the weights ρ_k's are defined in (3.8), and let $\hat{\tau}_\delta(a,b)$ be the new stopping time that applies $\tau(a,b)$ to the new observations after time $\delta_{\max}$.

Now whenever $\hat{\tau}_\delta(a,b)$ stops at time $n_0 + \delta_{\max}$, we know that $\tau(a,b)$ stops after applying it to n_0 observations $(X_{k,\delta_{\max}+1},\cdots,X_{k,\delta_{\max}+n_0})$ for each k. By the definition of the local CUSUM statistics in (2.5), we have

$$W_{k,n_0+\delta_{\max}} \geq \sum_{i=\delta_{\max}+1}^{\delta_{\max}+n_0} \log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq \rho_k b$$

for all $1 \leq k \leq m$. Hence,

$$\sum_{k=1}^K W_{k,n_0+\delta_{\max}} \mathbf{1}\{W_{k,n_0+\delta_{\max}} \geq \rho_k b\} \geq \sum_{k=1}^m \sum_{i=\delta_{\max}+1}^{\delta_{\max}+n_0} \log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq a,$$

where the last relation is from the definition of $\tau(a,b)$. This implies that the scheme $N_{hard}(a,b)$ must stop at time $n_0 + \delta_{\max}$, and possibly earlier. Thus

$$\mathbf{E}^{(\nu=1)}_{\delta_1,\ldots,\delta_K}(N_{hard}(a,b)) \ \leq \ \mathbf{E}^{(\nu=1)}_{\delta_1,\ldots,\delta_K}(\hat{\tau}_\delta(a,b)) = \delta_{\max} + \mathbf{E}^{(\nu=1)}_{\delta_1^*,\ldots,\delta_K^*}(\tau(a,b)),$$

where δ_k^* is the binary version of δ_k's defined in (5.10). To simplify the notation, denote by $\mathbf{E}^{(1)}$ the expectation when the change occurs at time $\nu = 1$ and the event affects the first m data streams immediately but does not affect the other remaining $K - m$ data streams. So it suffices to show that the stopping time $\tau(a, b)$ in (5.2) satisfies

$$\mathbf{E}^{(1)}(\tau(a,b)) \leq \max\left\{\frac{a}{\sum_{k=1}^{m} I(g_k, f_k)}, \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\right\} + O(\sqrt{b}) + O(1). \tag{5.3}$$

To prove (5.3), for $1 \leq k \leq m$, let

$$\begin{aligned} M_k &= \inf\Big\{n \geq 1 : \sum_{i=1}^{n} \log \frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq \rho_k b\Big\}, \\ \tau_k(M_k) &= \sup\Big\{n \geq 1 : \sum_{i=M_k+1}^{M_k+n} \log \frac{g_k(X_{k,i})}{f_k(X_{k,i})} \leq 0\Big\} \\ \hat{M} &= \max_{1 \leq k \leq m} \Big(M_k + \tau_k(M_k) + 1\Big) \\ t(\hat{M}) &= \inf\Big\{n \geq 1 : \sum_{i=\hat{M}+1}^{\hat{M}+n} \Big(\sum_{k=1}^{m} \log \frac{g_k(X_{k,i})}{f_k(X_{k,i})}\Big) \geq \max\{a - (\sum_{k=1}^{m} \rho_k) b, 0\}\Big\}. \end{aligned}$$

Combining these definitions with those of $\tau(a, b)$ in (5.2) yields that

$$\begin{aligned} \tau(a,b) &\leq \hat{M} + t(\hat{M}) = \max_{1 \leq k \leq m} \Big(M_k + \tau_k(M_k) + 1\Big) + t(\hat{M}) \\ &\leq \sum_{k=1}^{m} \tau_k(M_k) + 1 + t(\hat{M}) + \max_{1 \leq k \leq m} M_k. \end{aligned}$$

Hence, relation (5.3) holds if we can establish the following three relations:

$$\mathbf{E}^{(1)}\Big(\tau_k(M_k)\Big) = O(1) \qquad \text{for all } 1 \leq k \leq m; \tag{5.4}$$

$$\mathbf{E}^{(1)}\Big(t(\hat{M})\Big) \leq \max\left\{\frac{a}{\sum_{k=1}^{m} I(g_k, f_k)} - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}, 0\right\} + O(1); \tag{5.5}$$

$$\mathbf{E}^{(1)}\Big(\max_{1 \leq k \leq m} M_k\Big) \leq \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + O(\sqrt{b}) + O(1). \tag{5.6}$$

Relation (5.4) is well-known in renewal theory, e.g., Theorem D in Kiefer and Sacks (1963), since $\log\big(g_k(X)/f_k(X)\big)$ has positive mean and finite variance under $\mathbf{E}^{(1)}$ by our assumptions in (2.1) and (2.2).

For relation (5.5), by the definition of $t(\hat{M})$, when $a \leq (\sum_{k=1}^{m} \rho_k) b$, the threshold becomes 0 and thus $t(\hat{M}) = 0$. When $a \geq (\sum_{k=1}^{m} \rho_k) b$, the stopping time $t(\hat{M})$ is defined when a random walk exceeds the bound $a - (\sum_{k=1}^{m} \rho_k) b$, the application of standard renewal theory yields that

$$\mathbf{E}^{(1)}(t(\hat{M})) = \frac{a - (\sum_{k=1}^{m} \rho_k) b}{\sum_{k=1}^{m} I(g_k, f_k)} + O(1)$$

$$= \frac{a}{\sum_{k=1}^{m} I(g_k, f_k)} - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + O(1),$$

see, for example, (Siegmund, 1985, Ch. VIII). Here the second equation follows from the definition of ρ_k in (3.8) that

$$\frac{\sum_{k=1}^{m} \rho_k}{\sum_{k=1}^{m} I(g_k, f_k)} = \frac{1}{\sum_{k=1}^{K} I(g_k, f_k)}.$$

Thus relation (5.5) holds.

The proof of relation (5.6) is a little more complicated, but it can be done along the same line as that in Mei (2005). The key fact is that the choice of $b_k = \rho_k b$'s in (3.7)-(3.8) makes sure that the stopping times M_k's have roughly the same mean under $\mathbf{P}^{(1)}$. Specifically, by renewal theory and the assumptions of (f_k, g_k) in (2.1) and (2.2), under $\mathbf{P}^{(1)}$,

$$\mathbf{E}^{(1)}(M_k) = \frac{\rho_k b}{I(g_k, f_k)} + O(1) = \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + O(1)$$

and $\text{Var}^{(1)}(M_k) = O(b)$, as $b \to \infty$, see Siegmund (Siegmund, 1985, p. 171). Thus as $b \to \infty$,

$$\begin{aligned}
\Big(\mathbf{E}^{(1)}\Big|M_k - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\Big|\Big)^2 &\leq \mathbf{E}^{(1)}\Big(M_k - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\Big)^2 \\
&= \text{Var}^{(1)}(M_k) + \Big(\mathbf{E}^{(1)} M_k - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\Big)^2 \\
&\leq C_{1k} b,
\end{aligned}$$

where $C_{1k} > 0$ is a constant. Taking square root both sides, and noticing that $M_k = M_k(b)$ is an increasing function of $b \geq 0$, it is not difficult to show that for each $k = 1, \cdots, K$, there exists a constant $C_{2k} > 0$ so that

$$\mathbf{E}^{(1)}\Big|M_k - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\Big| \leq \max(C_{2k}, \sqrt{C_{1k}}\sqrt{b}),$$

for all $b > 0$.

Therefore,

$$\begin{aligned}
\mathbf{E}^{(1)}\Big(\max_{1 \leq k \leq m} M_k\Big) &= \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + \mathbf{E}^{(1)} \max_{1 \leq k \leq m} \Big(M_k - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\Big) \\
&\leq \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + \sum_{k=1}^{m} \mathbf{E}^{(1)}\Big|M_k - \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)}\Big| \\
&\leq \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + \sum_{k=1}^{m} \max(C_{2k}, \sqrt{C_{1k}}\sqrt{b}) \\
&\leq \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + C(\sqrt{b} + 1),
\end{aligned}$$

where the constant $C = \sum_{k=1}^{K} \max(C_{2k}, \sqrt{C_{1k}})$ does not depend on b. This proves relation (5.6). Therefore, relations (5.4)-(5.6) hold, and thus relation (4.3) holds for the hard-thresholding scheme $N_{hard}(a, b)$ in (3.2).

The proof for the soft-thresholding scheme $N_{soft}(a,b)$ in (3.3) is similar, except defining the stoping time $\tau(a,b)$ by

$$\tau(a,b) = \text{first } n \text{ such that } \sum_{i=1}^{n}\sum_{k=1}^{m}\log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq a + b\sum_{k=1}^{m}\rho_k \text{ and} \tag{5.7}$$

$$\sum_{i=1}^{n}\log\frac{g_k(X_{k,i})}{f_k(X_{k,i})} \geq \rho_k b \text{ for all } 1 \leq k \leq m, \tag{5.8}$$

instead of (5.2) and prove

$$\mathbf{E}^{(1)}(\tau(a,b)) \leq \frac{a}{\sum_{k=1}^{m} I(g_k, f_k)} + \frac{b}{\sum_{k=1}^{K} I(g_k, f_k)} + O(\sqrt{b}) + O(1). \tag{5.9}$$

by replacing the threshold $\max\{a - (\sum_{k=1}^{m}\rho_k)b, 0\}$ in the stopping time $t(\hat{M})$ by the threshold a. The remaining arguments are identical and thus omitted.

Now let us provide a sketch of the proof for part (iii) of Theorem 4.1 on the order-thresholding scheme $N_{order,r}(a)$ in (3.6) and the combined thresholding scheme $N_{comb,r}(a,b)$ in (3.5). Since $N_{order,r}(a)$ is a special case of $N_{comb,r}(a,b)$ with $b = 0$, it suffices to prove the theorem for $N_{comb,r}(a,b)$ in (3.5) with $b \geq 0$. Clearly relation (5.11) also holds for $N_{comb,r}(a,b)$ for any $b \geq 0$, because the "SUM" scheme $T_{\text{sum}}(a)$ again provides the lower bound for $N_{comb,r}(a,b)$.

It remains to show that relation (4.3) holds for $N_{comb,r}(a,b)$ with $b \geq 0$ in the scenario when the occurring event affects at most r data streams, i.e., when $\sum_{k=1}^{K} I\{\delta_k < \infty\} \leq r$. Without loss of generality, assume that the affected data streams are just the first m data streams with $m \leq r$. Recall that $U_{k,n} = W_{k,n} I\{W_{k,n} \geq \rho_k b\}$, and we order the $U_{k,n}$'s as $U_{(1),n} \geq \ldots \geq U_{(K),n}$, and $N_{comb,r}(a,b)$ stops if $\sum_{k=1}^{r} U_{(k),n} \geq a$. Note that if $m \leq r$,

$$\sum_{k=1}^{r} U_{(k),n} \;\; \geq \;\; \sum_{k=1}^{r} U_{k,n} \geq \sum_{k=1}^{m} U_{k,n},$$

since $U_{k,n} \geq 0$. Thus, if at some time n_0 we have $W_{k,n_0} \geq \rho_k b$ and $\sum_{k=1}^{m} W_{k,n_0} \geq a$ for $1 \leq k \leq m$ (i.e., for the first m data streams), then $N_{comb,r}(a,b)$ will also stop at time n_0 and possibly earlier. Hence, whenever $m \leq r$, the stopping time $\tau(a,b)$ in (5.2) also provides an upper bound on the detection delay of $N_{comb,r}(a,b)$. Thus the proposed combined thresholding scheme $N_{comb,r}(a,b)$ in (3.5) satisfies relation (4.3) whenever the occurring event affects at most r data streams. This completes the proof of the theorem. □

Proof of Lemma 4.1. Intuitively, only those affected sensors provide information to detect the occurring events, and the quickest possible way to detect the occurring event is when the event affects the sensors instantaneously. More rigorously, if we define

$$\delta_k^* = \begin{cases} 0, & \text{if } \delta_k \text{ is finite} \\ \infty, & \text{if } \delta_k = \infty \end{cases}, \tag{5.10}$$

then for any given scheme $T(\gamma)$,

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(T(\gamma)) \geq \inf_{\tau} \overline{\mathbf{E}}_{\delta_1^*,\ldots,\delta_K^*}(\tau),$$

where the infimum is taken over all possible schemes τ satisfying the false alarm constraint γ in (2.4). An alternative and possible better viewpoint is based on a time-shifting argument in which one imagines that at time n one observes the observations $X_{k,n+\delta_k}$ (instead of $X_{k,n}$) when δ_k is finite, and then applies $T(\gamma)$ to the new aligned observations.

Without loss of generality, assume that the first m data streams are affected abruptly and simultaneously by the event at unknown time ν, and other data streams are unaffected. That is, m out of K data streams are affected by the event, and $\delta_i^* = 0$ for $1 \leq i \leq m$, and $= \infty$ for $m+1 \leq i \leq K$. By (4.2), we have

$$J(\delta_1, \ldots, \delta_K) = J(\delta_1^*, \ldots, \delta_K^*) = \sum_{i=1}^{m} I(g_i, f_i).$$

In this case, we face the sequential change detection problem when the distribution of $(X_{1,n}, \cdots, X_{K,n})$ changes from $(f_1, \cdots, f_m, f_{m+1}, \cdots, f_K)$ to $(g_1, \cdots, g_m, f_{m+1}, \cdots, f_K)$. It is well-known (Lorden (1971)) that

$$\inf_{\tau} \overline{\mathbf{E}}_{\delta_1^*, \ldots, \delta_K^*}(\tau) \geq (1 + o(1)) \frac{\log \gamma}{\sum_{i=1}^{m} I(g_i, f_i)}.$$

subject to the false alarm constraint γ in (2.4) as $\gamma \to \infty$. Combining the above results yields relation (4.5), completing the proof of Lemma 4.1. □

Proof of Theorem 4.2: First, we will prove for any $a, b \geq 0$,

$$\mathbf{E}^{(\infty)}(N_{hard}(a, b)) \geq (1 + o(1)) \frac{e^a}{1 + a + \frac{a^2}{2!} + \cdots + \frac{a^{K-1}}{(K-1)!}}. \tag{5.11}$$

To prove (5.11), note that $N_{hard}(a, b)$ in (3.2) is increasing as a function of $b \geq 0$, and when $b = 0$, $N_{hard}(a, b = 0)$ reduces to the "SUM" scheme $T_{\text{sum}}(a)$ in (2.7). Hence, for any $b \geq 0$, $N_{hard}(a, b) \geq T_{\text{sum}}(a)$ and of course, $\mathbf{E}^{(\infty)}(N_{hard}(a, b)) \geq \mathbf{E}^{(\infty)}(T_{\text{sum}}(a))$. By Theorem 1 of Mei (2010), the "SUM" scheme $T_{\text{sum}}(a)$ satisfies relation (5.11), and so are the hard-thresholding schemes $N_{hard}(a, b)$ for all $b \geq 0$.

Theorem 4.2 follows at once from Theorem 4.1 and 5.11. In particular, the choice of a_γ in (4.6) follows from (5.11) and the fact that $1 + a + \frac{a^2}{2!} + \cdots + \frac{a^{K-1}}{(K-1)!} \sim \frac{a^{K-1}}{(K-1)!}$ if K is fixed and a goes to ∞. □

Proof of Theorem 4.3: Clearly, we can see for any fixed combination of (a, b), $\mathbf{E}^{(\infty)} N_{hard}(a, b)$ is smaller than $\mathbf{E}^{(\infty)} N_{soft}(a, b)$ or $\mathbf{E}^{(\infty)} N_{comb,r}(a, b)$. Therefore, it is sufficient to prove the choice of a in (4.8) could guarantee the hard-thresholding scheme $N_{hard}(a, b)$ satisfies false alarm constraint (2.4).

First, define $W_k^* = \lim_{n \to \infty} W_{k,n}$ as the limit of the CUSUM statistics, which has the following non-asymptotic result: for *any* $x > 0$, the tail probability

$$G(x) = \mathbf{P}^{(\infty)}(W_k^* > x) \leq e^{-x}, \tag{5.12}$$

see Appendix 2 on Page 245 of Siegmund (1985). It is clear that W_k^* are i.i.d. across different k. Now we define the log-moment generating function of the W_k^*'s

$$\psi(\theta) = \log \mathbf{E}^{(\infty)} \exp\{\theta W_k^* \mathbf{1}\{W_k^* \geq b/K\}\} \tag{5.13}$$

For any $x \geq 0$, by Chebyshev's inequality,

$$
\begin{aligned}
\mathbf{E}^{(\infty)}[N_{hard}(a,b)] &\geq x\mathbf{P}^{(\infty)}(N_{hard}(a,b) \geq x) \\
&= x\left[1 - \mathbf{P}^{(\infty)}(N_{hard}(a,b) < x)\right] \\
&= x\left[1 - \mathbf{P}^{(\infty)}(\sum_{k=1}^{K} W_{k,n}\mathbf{1}\{W_{k,n} \geq b_k\} \geq a) \text{ for some } 1 \leq n \leq x\right] \\
&\geq x\left[1 - x\mathbf{P}^{(\infty)}(\sum_{k=1}^{K} W_k^* \mathbf{1}\{W_k^* \geq b_k\} \geq a)\right], \\
&\geq x\left[1 - xe^{-\theta a}\mathbf{E}^{(\infty)}\exp(\theta \sum_{k=1}^{K} W_k^* \mathbf{1}\{W_k^* \geq b/K\})\right] \\
&= x\left[1 - x\exp\left(-\theta a + K\psi(\theta)\right)\right]. \qquad (5.14)
\end{aligned}
$$

Note that for any $u > 0$, the function $x(1 - xu)$ is maximized at $x = 1/(2u)$ with the maximum value $1/(4u)$. Therefore, we can get for any $0 < \theta < 1$,

$$
\mathbf{E}^{(\infty)}[N_{hard}(a,b)] \geq \frac{1}{4}\exp\left(\theta a - K\psi(\theta)\right). \qquad (5.15)
$$

By the definition of $\psi(\theta)$ in (5.13) and the tail probability W_k^* in (5.12), for all $0 < \theta < 1$,

$$
\begin{aligned}
\psi(\theta) &= \log[\mathbf{P}^{(\infty)}(W_k^* \leq b/K) - \int_{b/K}^{\infty} e^{\theta x} dG(x)] \\
&= \log[1 + (e^{\theta b/K} - 1)G(b) + \theta \int_{b/K}^{\infty} e^{\theta x} G(x)dx)] \\
&\leq \log[1 + (e^{\theta b/K} - 1)e^{-b/K} + \theta \int_{b/K}^{\infty} e^{\theta x} G(x)dx] \\
&\leq \log[1 + (e^{\theta b/K} - 1)e^{-b/K} + \theta \int_{b/K}^{\infty} e^{\theta x} e^{-x}dx] \\
&= \log\left(1 + \frac{1}{1-\theta}e^{-b(1-\theta)/K} - e^{-b/K}\right) \\
&\leq \frac{1}{1-\theta}e^{-b(1-\theta)/K} - e^{-b/K} \\
&\leq \frac{1}{1-\theta} - e^{-b/K} \qquad (5.16)
\end{aligned}
$$

where the second equation is based on the integration by parts. By (5.15) and (5.16), we have

$$
\mathbf{E}^{(\infty)} N_{hard}(a,b) \geq \frac{1}{4}\exp\left(\theta a - \frac{K}{1-\theta} + Ke^{-b/K}\right) \qquad (5.17)
$$

for all $0 < \theta < 1$. If $K < a$, by letting $\theta = 1 - \sqrt{K/a}$ yield

$$
\mathbf{E}^{(\infty)} N_{hard}(a,b) \geq \frac{1}{4}\exp\left((\sqrt{a} - \sqrt{K})^2 + Ke^{-b/K} - K\right) \qquad (5.18)
$$

Therefore a choice of

$$a = (\sqrt{\log(4\gamma) + K - Ke^{-b/K}} + \sqrt{K})^2 \tag{5.19}$$

will guarantee the hard-shrinkage scheme $N_{hard}(a, b)$ satisfies the false alarm constraint (2.4).

Note using the continuity of the soft-thresholding transformation function, a tighter bound for $N_{soft}(a, b)$ was derived for the soft-thresholding scheme in Liu et al. (2018), although they are asymptotically equivalent to those in Theorem 4.3 and Corollary 4.1 $N_{hard}(a, b)$ as the dimension K goes to ∞. □

Proof of Corollary 4.1: If $K = o(\log\gamma)$, the corresponding $a = a_\gamma = \log(4\gamma) + o(\log\gamma)$ will guarantee the false alarm constraint. Moreover, if m is fixed and $b = o(\log\gamma)$, the upper bound of detection delay in theorem 4.1 could be applied and yields

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a, b)) \leq (1 + o(1)) \left(\frac{\log\gamma}{mI}\right) + O(1), \tag{5.20}$$

which implies the first order detection efficiency will be kept as long as $b = o(\log\gamma)$.

If $K >> \log\gamma$, the corresponding $a = (1 + o(1))K$ will guarantee the false alarm constraint. Moreover, since m is fixed and $b = o(K^2)$, the upper bound of detection delay in theorem 4.1 could be applied and yields

$$\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a, b)) \leq (1 + o(1)) \left(\frac{K}{mI}\right) + O(1), \tag{5.21}$$

which completes the proof of corollary. □

ACKNOWLEDGEMENTS

This research is partially supported by NSF grants CMMI-1362876, DMS-1613258, and DMS-1830344.

REFERENCES

Appadwedula, S., Veeravalli, V. V., and Jones, D. (2005). Energy-Efficient Detection in Sensor Networks. *IEEE Journal on Selected Areas in Communications* 23: 693–702.

Banerjee, S. and Fellouris, G. (2016). Decentralized Sequential Change Detection with Ordered CUSUMs. *Proceedings of 2016 IEEE International Symposium on Information Theory (ISIT 2016)*, July 10-15, 2016, pp. 36–40. Barcelona, Spain.

Banerjee, T. and Veeravalli, V. V., (2015). Data-Efficient Quickest Change Detection in Sensor Networks. *IEEE Transactions on Signal Processing* 63: 3727–3735.

Bartroff, J. (2018). Multiple Hypothesis Tests Controlling Generalized Error Rates for Sequential Data. *Statistica Sinica* 28: 363–398.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Applications.* Englewood Cliffs, Prentice-Hall.

Candès, E. J. (2006). Modern Statistical Estimation via Oracle Inequalities. *Acta Numerica* 15: 257–325.

Chan, H. P. (2017). Optimal Sequential Detection in Multi-Stream Data. *The Annals of Statistics* 45: 2736–2763.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika* 81: 425–455.

Durrett, R. (1996). *Probability: Theory and Examples*. Second edition. Duxbury Press, Belmont, CA.

Fan, J. and Lin, S. K. (1998). Test of Significance When Data are Curves. *Journal of American Statistical Association* 93: 1007–1021.

Fuh, C.D. and Mei, Y. (2015). Quickest Change Detection and Kullback-Leibler Divergence for Two-State Hidden Markov Models. *IEEE Transactions on Signal Processing* 63: 4866–4878.

Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics*. Springer-Verlag, New York.

Gordon, L. and Pollak, M. (1994). An Efficient Sequential Nonparametric Scheme for Detecting a Change of Distribution. *The Annals of Statistics* 22: 763–804.

Kiefer, J. and Sacks, J. (1963). Asymptotically Optimum Sequential Inference and Design. *The Annals of Mathematical Statistics* 34: 705–750.

Kulldorff, M. (2001). Prospective Time-Periodic Geographic Disease Surveillance Using a Scan Statistic, *Journal of Royal Statistical Society, Series A* 164: 61–72.

Lai, T. L. (1995). Sequential Change-Point Detection in Quality Control and Dynamical Systems (with Discussion). *Journal of Royal Statistical Society, Series B* 57: 613–658.

Lai, T. L. (2001). Sequential Analysis: Some Classical Problems and New Challenges. *Statistica Sinica* 11: 303–408.

Lévy-Leduc, C. and Roueff, F. (2009). Detection and Localization of Change-points in High-Dimensional Network Traffic Data. *The Annals of Applied Statistics* 3: 637–662.

Liu, K., Mei, Y. and Shi, J. (2015). An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring. *Technometrics* 57: 305–319.

Liu, K., Zhang, R. and Mei, Y. (2018). Scalable Sum-Shrinkage Schemes for Distributed Monitoring Large-Scale Data Streams. *Statistica Sinica* (In press).

Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *The Annals of Mathematical Statistics* 42: 1897–1908.

Lorden, G. and Pollak, M. (2008). Sequential Change-Point Detection Procedures That Are Nearly Optimal and Computationally Simple. *Sequential Analysis* 27:476-512.

Mei, Y. (2005). Information Bounds and Quickest Change Detection in Decentralized Decision Systems. *IEEE Transactions on Information Theory* 51:2669–2681.

Mei, Y. (2010). Efficient Scalable Schemes for Monitoring a Large Number of Data Streams. *Biometrika* 97: 419-433.

Mei, Y. (2011). Quickest Detection in Censoring Sensor Networks. In *IEEE International Symposium on Information Theory (ISIT)*, Aug. 2011, pp. 2148–2152.

Montgomery, D. C. (1991). *Introduction to Statistical Quality Control* (2nd edition). Wiley, New York.

Moustakides, G. V. (1986). Optimal Stopping Times for Detecting Changes in Distributions. *The Annals of Statistics* 14:1379–1387.

Neyman, J. (1937). Smooth Test for Goodness-of-Fit. *Skand. Aktuarietidskr* 20: 149–199.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika* 41: 100–115.

Pollak, M. (1985). Optimal Detection of a Change in Distribution. *The Annals of Statistics* 13: 206–227.

Pollak, M. (1987). Average Run Lengths of an Optimal Method of Detecting a Change in Distribution. *The Annals of Statistics* 15: 749–779.

Poor, H. V. and Hadjiliadis, O. (2009). *Quickest Detection*. Cambridge University Press, New York, 2009.

Rago, C., Willett, P. and Bar-Shalom, Y. (1996). Censoring Sensors: A Low-Communication-Rate Scheme for Distributed Detection. *IEEE Transactions on Aerospace and Electronic Systems* 32: 554–568.

Roberts, S. W. (1966). A Comparison of Some Control Chart Procedures. *Technometrics* 8: 411–430.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. D Van Norstrand, New York. Preprinted by ASQC Quality Press, Wisconsin, 1980.

Shiryaev, A. N. (1963). On Optimum Methods in Quickest Detection Problems. *Theory of Probability & Its Applications* 8: 22–46.

Siegmund, D. (1985): *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

Tartakovsky, A., Nikiforov, I. and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman and Hall/CRC.

Tartakovsky, A. G., Rozovskiia, B. L., Blazeka, R. B. and Kim, H. (2006). Detection of Intrusions in Information Systems by Sequential Change-Point Methods (with discussions). *Statistical Methodology* 3: 252–340.

Tartakovsky, A. G. and Veeravalli, V. V. (2004). Change-Point Detection in Multichannel and Distributed Systems, *Applications of Sequential Methodologies*, N. Mukhopadhyay, S. Datta and S. Chattopadhyay, eds., pp. 339–370, New York: Marcel Dekker.

Tay, W. P., Tsitsiklis, J. N. and Win, M. Z. (2007). Asymptotic Performance of a Censoring Sensor Network. *IEEE Transactions on Information Theory* 53: 4191–4209.

Veeravalli, V. V. (2001). Decentralized Quickest Change Detection. *IEEE Transactions on Information Theory* 47: 1657–1665.

Wang, Y. and Mei, Y. (2015). Large-Scale Multi-Stream Quickest Change Detection via Shrinkage Post-Change Estimation, *IEEE Transactions on Information Theory* 61: 6926–6938.

Wang, Y., Mei, Y., and Paynabar, K. (2018). Thresholded Multivariate Principal Component Analysis for Phase I Multichannel Profile Monitoring, *Technometrics* 60: 360–372.

Xian, X., Wang, A., and Liu, K. (2018). A Nonparametric Adaptive Sampling Strategy for Online Monitoring of Big Data Streams. *Technometrics* 60:14–25.

Xie, Y., Huang, J. and Willett, R. (2013). Changepoint Detection for High-Dimensional Time Series with Missing Data. *IEEE Journal of Selected Topics in Signal Processing* 7: 12–27.

Xie, Y. and Siegmund, D. (2013). Sequential Multi-Sensor Change-Point Detection. *The Annals of Statistics* 41: 670–692.

Zhang, R., Mei, Y., and Shi, J. (2018). Wavelet-Based Profile Monitoring Using Order-Thresholding Recursive CUSUM Schemes. Accepted as a book chapter in *New Frontiers in Biostatistics and Bioinformatics*, Y. Zhao, and D.G. Chen, eds., Switzerland: Springer.