Quickest Detection in Censoring Sensor Networks

Yajun Mei
School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, Georgia 30332–0225
Email: ymei@isye.gatech.edu

Abstract—**The quickest change detection problem is studied in a general context of monitoring a large number of data streams in sensor networks when the "trigger event" may affect different sensors differently. In particular, the occurring event could have an immediate or delayed impact on some unknown, but not necessarily all, sensors. Motivated by censoring sensor networks, scalable detection schemes are developed based on the sum of those local CUSUM statistics that are "large" under either hard thresholding or top-r thresholding rules or both. The proposed schemes are shown to possess certain asymptotic optimality properties.**

I. 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 [15].

In this paper, we consider a novel and general scenario of quickest detection problems when the "trigger event" may affect different sensors differently. A naive approach is to monitor each local sensor individually. Unfortunately, the local monitoring approach does not take advantage of global information, and may lead to large detection delays if several data streams can provide information about the occurring event. More importantly, even if the local false alarm rate is well controlled at each local sensor, the system-wise global false alarm rate can be severe when the number of sensors (or data streams) is large.

The main purpose of this paper is to illustrate how to combine many local monitoring schemes together to produce a single global monitoring scheme, so that the system can detect the occurring event as quickly as possible subject to a system-wise global false alarm constraint. Our proposed methodology was motivated by censoring sensor network, which was introduced by Rago et al. [10] and later by Appadwedula et al. [1] and by Tay et al. [14]. 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) are generally cheap and affordable, communication between remote sensors and fusion center are 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.

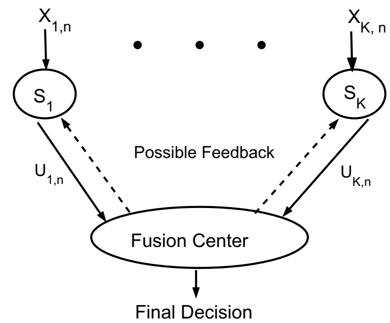

Fig. 1. A widely used configuration of censoring sensor networks.

This consideration motivates us to propose to raise a global alarm based on the sum of those local detection statistics (e.g., local CUSUM statistics) that are "large" under either hard thresholding rule or top-r thresholding or both. It is worth pointing out that a well-known view in the standard off-line statistical inference literature is the necessity of thresholding for high-dimensional data in order to improve power or efficiency, see Candes [3] and the references there. However, our application of thresholding rules to sequential change-point detection seems to be new.

The remainder of this article is organized as follows. In Section II 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. Section III defines our proposed methodologies, and Section IV discusses how to choose parameters in the proposed methodologies. Section V presents an asymptotic optimality theory and Section VI reports numerical Monte Carlo simulation results.

II. Problem Formulation and Existing Methodologies

Suppose that in a network system, 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

978-1-4577-0595-3/11/$26.00 ©2011 IEEE

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 of 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 detection scheme, so that the system can detect the occurring event as quickly as possible. Denote 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 denote 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 $\mathbf{P}^{(\infty)}$.

As in the classical quickest change detection problems, a detection scheme can be defined as a stopping time T, and the interpretation of N is that, when $N = n$, we stop at time n and declare that a change has occurred. Our problem can then be formulated as to find a stopping time T such that the detection delay

$$\begin{aligned} & \overline{\mathbf{E}}_{\delta_1,\delta_2,\ldots,\delta_K}(T) \\ = & \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbf{E}^{(\nu)}_{\delta_1,\delta_2,\ldots,\delta_K} \left((T - \nu + 1)^+ \middle| \mathcal{F}_{\nu-1} \right) \end{aligned}$$

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

$$\mathbf{E}^{(\infty)}(T) \geq \gamma, \tag{1}$$

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

In our asymptotic theory below, we will make some simplified assumptions about the post-change hypothesis set Δ on the delay effects $(\delta_1, \delta_2, \ldots, \delta_k)$. First, we assume that at least one of δ_k's is finite if a change occurs, since otherwise the change is undetectable. Second, without loss of generality, the minimum delay effect is assumed to be 0, as we can always define a new change-point $\nu' = \nu + \min_{1 \leq k \leq K} \delta_k$. Finally, we assume that the finite delay affects δ_k's are not too large as compared to $\log \gamma$, the order of typical detection delays subject to the false alarm constraint, since otherwise such sensor will not pay a role in quickest detection. To summarize, the delay effects $(\delta_1, \delta_2, \ldots, \delta_k)$ are assumed to belong to

$$\begin{aligned} \Delta \;\; = \;\; & \big\{ (\delta_1, \ldots, \delta_K) : \delta_k\text{'s either} = \infty \text{ or satisfy} \\ & 0 \leq \delta_k << \log \gamma, \text{ and } \min_{1 \leq k \leq K} \delta_k = 0 \big\}, \end{aligned}$$

as γ goes to ∞. Here $x(a) << y(a)$ implies that $x(a)/y(a) \to 0$ as $a \to \infty$. In our asymptotic theory, we assume that the delay effects set Δ in (2) and the unknown change-point ν are separated, i.e., the set Δ does not depend on the unknown change-point ν. Under our setting, detecting the unknown change-point ν is of primary interest, and the delay effects δ_k's are nuisance parameters.

When $K = 1$ or when monitoring a single local data stream, say, the kth data stream, such a problem has been well studied in the sequential change-point detection literature, see [2], [4], [5], [7], [8], [9], [11]. One efficient local scheme 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

$$\begin{aligned} W_{k,n} &= \max \Big\{ 0, \max_{1 \leq \nu \leq n} \sum_{i=\nu}^{n} \log \frac{g_k(X_{k,i})}{f_k(X_{k,i})} \Big\} \\ &= \max \Big(W_{k,n-1} + \log \frac{g_k(X_{k,n})}{f_k(X_{k,n})}, \; 0 \Big). \end{aligned} \tag{2}$$

Below we will develop global schemes based on the local CUSUM statistics $W_{k,n}$ in (2), although the ideas can be easily extended to other local detection statistics such as Shiryeav-Roberts statistics or scan statistics.

Now let us go back our global monitoring problem when K is moderately large, and it is known that the generalized likelihood ratio based methods are computationally infeasible, see [6]. Thus it is natural to combine the local detection schemes to develop efficient global schemes, 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 [13]. Mathematically, the "MAX" scheme raises a global alarm at time

$$T_{\max}(c) = \inf\{ n \geq 1 : \max_{1 \leq k \leq K} W_{k,n} \geq c \}, \tag{3}$$

($= \infty$ if such n does not exist) where $c > 0$ is a pre-specified constant chosen to satisfy the false alarm constraint (1). The second approach is the "SUM" scheme, proposed in Mei [6], 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), and then one will raise an alarm at the global level at time

$$T_{\mathrm{sum}}(d) = \inf\{ n \geq 1 : \sum_{k=1}^{K} W_{k,n} \geq d \}, \tag{4}$$

where the constant $d > 0$ is some suitably chosen constant. Intuitively, the "MAX" scheme $T_{\max}(c)$ in (3) works better when one or very few data streams are affected, whereas the "SUM" scheme $T_{\mathrm{sum}}(d)$ in (4) works better when many data streams are affected, and numerical simulations in Mei [6] indeed verified this intuition.

III. Our Proposed Methodology

In this section, we propose to develop global monitoring schemes via hard thresholding or top-r thresholding or both. It turns out that the proposed schemes integrate the existing "MAX" and "SUM" schemes, and offer new methodologies for online monitoring a large number of data streams.

A. Hard Thresholding Schemes

In 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, and the corresponding scheme will be called hard thresholding schemes.

Specifically, at time n, each local sensor calculates its local CUSUM statistic $W_{k,n}$ recursively as in (2), and then sends the following sensor message $U_{k,n}$ to the fusion center:

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

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 from the sensors, the fusion center then raises an alarm at the global level at first time n such that

$$\sum_{\text{received}} U_{k,n} \geq a,$$

where the detection threshold $a > 0$ is a pre-specified constant.

So far we simply follow our intuition without discussing how to choose the local censoring parameters b_k's, especially when the data streams are nonhomogeneous. It turns out that a "good" choice of the b_k's is $b_k = \rho_k b$ for $k = 1, \ldots, K$ for some constant $b \geq 0$ where

$$\rho_k = \frac{I(g_k, f_k)}{\sum_{k=1}^{K} I(g_k, f_k)} \tag{5}$$

can be thought of as the weight of the k-th data stream in the overall final decision, and $I(g_k, f_k)$ is the Kullback-Leibler information number defined by

$$I(g_k, f_k) = \int \log \frac{g_k(x)}{f_k(x)} g_k(x) d\mu(x). \tag{6}$$

With this choice of b_k's, the proposed hard thresholding scheme will raise a global alarm at time

$$N_{hard}(a, b) = \inf\{n : \sum_{k=1}^{K} W_{k,n} I\{W_{k,n} \geq \rho_k b\} \geq a\}, \tag{7}$$

where $a > 0$ and $b \geq 0$ are two suitably chosen constants.

Evidently, if the threshold parameter $b = 0$, then the hard thresholding scheme $N_{hard}(a, b)$ in (7) becomes the "SUM" scheme $T_{\text{sum}}(d)$ in (4) since the $W_{k,n}$'s are always non-negative by the definition in (2). On the other hand, if the threshold parameter b is very large, say $b \geq a / \min_{1 \leq k \leq K} \rho_k$, then the hard thresholding scheme $N_{hard}(a, b)$ in (7) becomes the "MAX" scheme $T_{\max}(c)$ in (3) with $c = \min_k(\rho_k b)$. Therefore, the family of schemes $N_{hard}(a, b)$ is actually a large family that includes both "MAX" and "SUM" schemes.

B. Top-r Thresholding Schemes

Note that the hard thresholding is in parallel to Type I censoring in engineering, statistics, and medical research, particularly reliability and survival analysis, in which an experiment has a set number of subjects or items and stops the experiment at a pre-determined time. Another type of censoring is the so-called Type II censoring in which one stops the experiment when a predetermined number of subjects are observed to have failed. This inspires us to construct the detection schemes based on the largest r CUSUM statistics.

To be more concrete, at time n, order the K local CUSUM statistics $W_{1,n}, \ldots, W_{K,n}$ from largest to smallest: $W_{(1),n} \geq W_{(2),n} \geq \ldots \geq W_{(K),n}$, and then the top-$r$ thresholding scheme raises a global alarm at time

$$N_{top,r}(a) = \inf\{n : \sum_{k=1}^{r} W_{(k),n} \geq a\}. \tag{8}$$

Obviously, the top-r thresholding scheme $N_{top,r}(a)$ becomes the "MAX" scheme $T_{\max}(c)$ when $r = 1$, and becomes the "SUM" scheme $T_{\text{sum}}(d)$ when $r = K$.

In practice, the top-r thresholding scheme $N_{top,r}(a)$ is useful in applications when one may have a prior knowledge that at most r out of K data streams will be affected by the occurring event. This is because in such scenario, intuitively (at most) r local CUSUM statistics will be significantly large to provide information to the occurring event.

Note that the top-r thresholding schemes can also be thought of hard-thresholding rules where the local censoring parameters are data-driven and adaptive over time n. Specifically, let $w_{r,n}$ denote the r-th largest statistics, i.e., $w_{r,n} = W_{(r),n}$, and then one raises an alarm at the global level at the time

$$N^*_{top,r}(a) = \inf\{n : \sum_{k=1}^{K} W_{k,n} I\{W_{k,n} \geq w_{r,n}\} \geq a\}.$$

Rigorously speaking, we have $N_{top,r}(a) \leq N^*_{top,r}(a)$, and they are equivalent only when the $W_{k,n}$'s are non-arithmetic, since they can be different when more than one of $W_{k,n}$'s is equal to $w_{r,n}$. Fortunately, both $N_{top,r}(a)$ and $N^*_{top,r}(a)$ possess similar asymptotic optimality properties.

C. The Combined Thresholding Schemes

For the purpose of applications in censoring sensor networks in Figure 1, one may combine the above-mentioned two thresholding methods together: use the hard thresholding at the local sensor level, and then adopt the top-r thresholding rule at the fusion center level to detect the event when one has a prior knowledge that the occurring event affects at most r sensors. Specifically, at time n, the sensor message sent by the k-th local sensor is defined by $U_{k,n} = W_{k,n} I\{W_{k,n} \geq \rho_k b\}$. The fusion center then orders all sensor messages $U_{k,n}$'s as $U_{(1),n} \geq \ldots \geq U_{(K),n}$, and raises an alarm if the sum of the r largest $U_{k,n}$'s is too large. Mathematically, the combined thresholding scheme raises an alarm at the global level at time

$$N_{comb,r}(a, b) = \inf\{n : \sum_{k=1}^{r} U_{(k),n} \geq a\}. \tag{9}$$

Alternatively, let $u_{r,n}$ be the r-th largest value among the K sensor messages $U_{k,n}$'s, and then another version of the combined thresholding scheme can be defined by

$$N^*_{comb,r}(a,b) = \inf\Big\{n : \sum_{k=1}^{K}\big[W_{k,n} I\{W_{k,n} \geq \rho_k b\} \times I\{W_{k,n} \geq u_{r,n}\}\big] \geq a\Big\}.$$

Again, $N_{comb,r}(a,b) \leq N^*_{comb,r}(a,b)$, but they are equivalent in non-arithmetic cases and are generally asymptotically equivalent otherwise. Evidently, the family of the combined thresholding schemes contains two censoring parameters, b and r, and it includes the families of both hard-thresholding and top-r thresholding schemes.

IV. Choice Of Thresholding Parameters

Compared to the existing "MAX" or "SUM" schemes, our proposed thresholding schemes include two new thresholding parameters: b for the hard-thresholding schemes and r for the top-r thresholding schemes (and both r and b for the combined thresholding schemes). It is natural to ask how to choose these two thresholding parameters in practice?

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 (7).

The choice of 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 [12]. 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 $N_{comb,r}(a,b)$ (or $N^*_{comb,r}(a,b)$) 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 (9) 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.

V. Asymptotic Optimality Theory

The following theorem states asymptotic optimality properties of our proposed thresholding schemes, as the false alarm constraint γ in (1) goes to ∞.

Theorem 1. *For a given K and for any $b \geq 0$, with the choice of*

$$a = a_\gamma = \log\gamma + (K - 1 + o(1))\log\log\gamma, \quad (10)$$

the proposed hard-threholding scheme $N_{hard}(a,b)$ satisfies the false alarm constraint (1) and asymptotically minimizes $\overline{\mathbf{E}}_{\delta_1,\ldots,\delta_K}(N_{hard}(a,b))$ (up to the first-order) for each and every post-change hypothesis $(\delta_1,\ldots,\delta_K) \in \Delta$ subject to the false alarm constraint (1), as γ in (1) goes to ∞. The conclusion also hold if $N_{hard}(a,b)$ is replaced by either the top-r thresholding scheme $N_{top,r}$ in (8) or the combined thresholding scheme $N_{comb,r}(a,b)$ in (9) when the occurring event affects at most r data streams, i.e., when $(\delta_1,\ldots,\delta_K) \in \Delta$ satisfies $\sum_{k=1}^{K} I\{\delta_k < \infty\} \leq r$.

Proof: Relation (10) follows from Theorem 1 of [6] and the fact that $N_{hard}(a,b) \geq N_{hard}(a, b=0) = T_{sum}(a)$. Asymptotic optimality properties can be derived by comparing the proposed schemes with the optimal CUSUM procedure that minimizes the detection delay for a given post-change hypothesis $(\delta_1,\ldots,\delta_K) \in \Delta$. The details are standard and thus omitted due to page limits. ■

VI. Numerical Simulations

Let us present a numerical simulation study in this section to illustrate the usefulness of the proposed schemes. Suppose that there are $K = 100$ independent and identical sensors in a system, and the observations at each sensor are i.i.d. with mean 0 and variance 1 before the change and with mean 0.5 and variance 1 after the possible change. Furthermore, we also assume that the change is instantaneous if a sensor is affected, i.e., the delay effect $\delta_k = 0$ or ∞.

For the purpose of comparison, we conduct numerical simulations for five families of detection schemes:

- the "MAX" scheme $T_{\max}(c)$ in (3),
- the "SUM" scheme $T_{\text{sum}}(d)$ in (4),
- the hard thresholding scheme $N_{hard}(a,b)$ in (7),
- the top-r scheme $N_{top,r}(a)$ in (8) with $r = 10$,
- the combined scheme $N_{comb,r}(a,b)$ in (9) with $r = 10$.

For each family of schemes $N_{hard}(a,b)$ and $N_{comb,r=10}(a,b)$, we further consider three specific schemes, depending on the

Fig. 2. Detection delays with $K = 100$ identical sensors

γ	Detection Scheme	# sensors affected 80	20	10	5	1
10^4	$T_{\mathrm{sum}}(d = 111.04)$	7.29 ± 0.02	20.1 ± 0.1	33.4 ± 0.2	55.2 ± 0.4	191.6 ± 2.1
	$N_{hard}(a = 106.38, b = 50)$	7.29 ± 0.02	20.2 ± 0.1	33.8 ± 0.2	56.1 ± 0.5	195.5 ± 2.1
	$N_{hard}(a = 62.26, b = 230.26)$	9.22 ± 0.03	19.7 ± 0.1	31.9 ± 0.2	53.7 ± 0.4	191.6 ± 2.1
	$N_{hard}(a = 29.70, b = 460.52)$	14.17 ± 0.05	21.9 ± 0.1	29.9 ± 0.2	43.3 ± 0.3	152.6 ± 1.7
	$T_{\max}(c = 11.12)$	32.74 ± 0.15	39.9 ± 0.2	45.2 ± 0.3	52.3 ± 0.4	85.5 ± 1.0
	Schemes $N_{comb,r}(a, b)$ in (9)					
	$N_{top,r=10}(a = 46.55)$	13.41 ± 0.04	20.8 ± 0.1	28.6 ± 0.2	41.8 ± 0.3	124.2 ± 1.4
	$N_{comb,r=10}(a = 46.55, b = 50)$	13.41 ± 0.04	20.8 ± 0.1	28.6 ± 0.2	41.8 ± 0.3	124.2 ± 1.4
	$N_{comb,r=10}(a = 46.53, b = 230.26)$	13.41 ± 0.04	20.8 ± 0.1	28.6 ± 0.2	42.3 ± 0.3	128.0 ± 1.4
	$N_{comb,r=10}(a = 29.70, b = 460.52)$	14.17 ± 0.04	21.9 ± 0.2	29.9 ± 0.2	43.4 ± 0.3	152.6 ± 1.8

value of the hard-thresholding parameter b: (i) $b = K/2$, (ii) $b = -\log(0.1)*K = 2.3026K$ and (iii) $b = -\log(0.01)*K = 4.6052K$. In the context of censoring sensor networks, the choices of these values will guarantee that when no event occurs, on average at most $\eta = \exp(-b/K) = 60.7\%, 10\%$, and 1% of $K = 100$ homogeneous sensors will transmit messages at any given time, respectively. Hence, in our simulation study, there are a total of nine specific monitoring schemes.

In our simulations we consider several post-change scenarios, depending on how many data streams are affected. The results are summarized in Table 1, which is based on 10^3 Monte Carlo simulations. From the table, we can see that when $5 \sim 10$ sensors are affected, the best scheme among these nine schemes belong to the family of schemes $N^*_{r=10}(a, b)$, which is designed to detect the scenario when 10 sensors are affected by the event. Also for each given scheme, the fewer affected sensors we have, the larger detection delay it will have.

It is worth mentioning that for the family of the hard-thresholding schemes $N_{hard}(a, b)$ in (7), a larger censoring threshold value b actually leads to a smaller detection delay when only a few (between 1 and 5) of data streams are affected. In other words, a larger censoring threshold value b in $N_{hard}(a, b)$ may actually be necessary for efficient detection when the affected data streams are sparse.

A surprising and possibly counter-intuitive result in Table 1 is the effect of not so large values of hard-thresholding parameter b in finite sample simulations. For example, the performances of the "SUM" scheme $T_{\mathrm{sum}}(d)$ and the hard thresholding scheme $N_{hard}(a, b = 50)$ are similar in view of sampling errors, and so are $N_{top,r=10}(a)$ and $N_{comb,r=10}(a, b = 50)$. That is, for $N_{hard}(a, b)$ or $N_{comb,r}(a, b)$, the schemes with $b = 0$ or $b = 50$ have similar performances, even though $b = 50$ implies that the scheme only requires $\exp(-b/K) = \exp(-0.5) = 60.7\%$ of 100 sensors to transmit information to the fusion center at any given time when no event occurs.

It is also interesting to see that the performances of $N_{hard}(a, b)$ and $N_{comb,r=10}(a, b)$ are different for small b, but are identical when $b = 460.52$ is large. Intuitively, the stopping time $N_{comb,r}(a, b)$ is decreasing as a function of r, and thus we should have $N_{hard}(a, b) \leq N_{comb,r=10}(a, b)$ even when $b = 460.52$. So one may wonder why our numerical simulations lead to identical results? One explanation is that with such a choice of $b = 460.52$, when no event occurs, on average there is at most 1 non-zero sensor messages 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.

In summary, from the performance viewpoint, using one of hard-thresholding and top-r thresholding approaches may be sufficient in certain applications, especially when the false alarm constraint γ in (1) is only moderately large.

REFERENCES

[1] APPADWEDULA, S., VEERAVALLI, V. V., and JONES, D. (2005). Energy-efficient detection in sensor networks. *IEEE J. Sel. Areas Commun.*, *23*, 693–702.

[2] BASSEVILLE, M. and NIKIFOROV, I. (1993). *Detection of Abrupt Changes: Theory and Applications.* Englewood Cliffs, Prentice-Hall.

[3] CANDS, E. J. (2006). Modern statistical estimation via oracle inequalities. Acta Numerica, *15*, 257–325.

[4] LAI, T. L. (2001). Sequential analysis: some classical problems and new challenges. *Statist. Sinica* **11** 303–408.

[5] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1897–1908.

[6] MEI, Y. (2010). Efficient scalable schemes for moniotoring a large number of data streams. *Biometrika*. **97** 419–433.

[7] MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387.

[8] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

[9] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227.

[10] RAGO, C., WILLETT, P., and BAR-SHALOM, Y. (1996). Censoring sensors: A low-communication-rate scheme for distributed detection. *IEEE Trans. Aerosp. Electon. Syst.*, *32*, 554–568.

[11] SHIRYAEV, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8** 22–46.

[12] SIEGMUND, D. (1985): *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

[13] TARTAKOVSKYA, 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.

[14] TAY, W.-P., TSITSIKLIS, J. N. and WIN, M. Z. (2007). Asymptotic performance of a censoring sensor network. *IEEE Trans. Inform. Theory* **53** 4191–4209.

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