Quickest Change Detection with Observation Scheduling

Xiaoqiang Ren*, Karl H. Johansson† and Ling Shi*

***Abstract*—The quickest change detection problem is to detect an abrupt change event as quickly as possible subject to constraints on false detection. Unlike the classical problem, where the decision maker can access only one sequence of observations, in this paper, the decision maker chooses one of two different sequences of observations at each time instant. The information quality and sampling cost of the two sequences of observations are different. We present an asymptotically optimal joint design of observation scheduling policy and stopping time such that the detection delay is minimized subject to constraints on both average run length to false alarm (ARLFA) and average cost per sample. The observation scheduling policy has a threshold structure and the detection scheme is a variant of the cumulative sum test where the detection statistic stochastically crosses the threshold that is used to switch observation modes. We further study the decentralized case in a multi-channel setting. We show that if each sensor uses the proposed observation scheduling policy locally and the fusion center uses the N_{sum} algorithm, by which the center declares the change when the sum of the sensors' local detection statistics crosses a certain threshold, the detection delay is asymptotically minimized for any possible combination of the affected sensors subject to constraints on both global ARLFA and average cost per sample at each sensor node. Numerical examples are given to illustrate the main results.**

Keywords: quickest change detection, minimax, observation scheduling, CuSum, sensor networks, multi-channel.

I. Introduction

Motivations: The quickest change detection with observation scheduling arises when multiple data streams about the monitored target are available but not all of those can be accessed at each time instant by the decision maker. In addition to determining the stopping time when the change event is declared, the decision maker needs to enhance the information quality of the observations taken using appropriate scheduling policy. This problem can be motivated by the following two examples. The first one is the detection with controlled sensing [1]–[6], where the detection system can adaptively control the information quality of observations used for decision making. For example, in clinical trials [4], patients may respond with different information about a drug treatment. For ethical reasons, the patients should be carefully and adaptively chosen over the period of a trial (usually several months or years), and the number of patients involved should be as small as possible. The second motivating example is the problem of detection with sensor networks [7], [8]. Consider the example of anomaly detection with different types of sensors, say proximity sensors and cameras. The sensors communicate with the central decision maker via wired or wireless channels. Obviously, the information quality of observations by proximity sensors and cameras are different. The cost of each sampling, including sensing energy consumption, bandwidth required to transmit, the computation load of data processing, of the observations by these two types of sensors is also different. Due to cost constraint, the decision maker thus needs to schedule its sampling attention. Note that even if there are only one type of sensors, the decision maker still may need to decide how many or which group of sensors to be activated.

The classical quickest change detection has been extensively studied, see the textbooks [9], [10]. In the classical problem formulation, the information quality of the observations is fixed and the decision maker only needs to determine the optimal stopping time. Recently, the quickest change detection with sampling constraints has attracted attention [1], [2], [11]–[13], where the decision maker needs to decide whether or not to sample at each time instant. The most relevant work is [1], where the authors proposed an asymptotically optimal data efficient CuSum algorithm called DE-CuSum algorithm. The DE-CuSum algorithm was studied in the sensor networks and multi-channel setting by the same authors in [2], [13]. Inspired by [1], [2], we consider the setting of observation scheduling, where, instead of sampling or not, the decision maker has to choose among multiple available observations. Take the intruder detection problem for example, at each time instant the decision maker has to decide to use the cheap proximity sensors or to activate the expensive cameras. Premkumar and Kumar [14] studied the optimal sleep–wake scheduling policy of wireless sensor networks for quickest change detection. The problem formulation is in the Bayesian setting [15], and since the sensors are homogenous, the observations available by switching on/off some sensor nodes have special structures. Our problem is studied in the minimax setting [16], [17] and the different data streams available have quite general characterizations. The CuSum algorithm with adaptive observations was also studied in the vehicle routing setting [18], where to detect anomalies as quickly as possible, a vehicle collect observations from adaptively chosen regions based on the likelihood of regional anomalies. The problem is fundamentally different from ours in the sense that it does

The work by X. Ren and L. Shi was supported by an RGC General Research Fund 16209114. The work by K. H. Johansson was supported in part by the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, and the Swedish Research Council.

*: Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong. Emails: xren@connect.ust.hk, eesling@ust.hk.

†: ACCESS Linnaeus Center and School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden. Email: kallej@ee.kth.se.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

not explicitly take the cost of observations into account in the problem formulation.

Contributions: In this paper, we formulate the quickest change detection with observation scheduling in the minimax setting. We assume that there exist two sequences of observations with different information about the underlying state and different cost for the decision maker, At each time instant, the decision maker needs to select one of these two sequences of observations. Due to the average cost per sample constraint, the observation scheduling policy needs to be carefully designed apart from the stopping time. We propose an algorithm wherein the observation scheduling policy has a threshold structure with respect to the detection statistic. The detection procedure is a variant of the cumulative sum (CuSum) algorithm [19] and a generalization of the DE-CuSum algorithm proposed in [1], where the detection statistic *stochastically* crosses that threshold used for the scheduling policy. This observation scheduling CuSum (OS-CuSum) algorithm is proved to asymptotically minimize the detection delay subject to constraints on both the average run length to false alarm (ARLFA) and average cost per sample (Theorem 2). It should be pointed out that though we assume that the decision maker chooses one of two sequences of observations at each time instant, the OS-CuSum algorithm can be easily generalized to scenarios where the decision maker can choose $m > 1$ out of $n > m$ sequences of observations. This is because any m observations taken at the same time can be treated as a single random vector, and the OS–CuSum algorithm can be easily generalized and remains asymptotically optimal in scenarios where the decision maker can take one of n sequences of observations (Remark 6).

We also study the quickest change detection with observation scheduling in the multi-channel setting [20], where the decision whether or not to stop is made by a fusion center based on multiple data streams collected by different sensor nodes. The sensor nodes affected by the change event is unknown. We assume that each sensor has different sensing modalities and needs to carefully schedule its sensing actions. We propose a multi-channel OS-CuSum (MOS-CuSum) algorithm: each sensor uses the observation scheduling policy locally and the fusion center uses the N_{sum} detection procedure proposed in [20], where the detection system stops when the sum of all the sensors' local detection statistic crosses a certain threshold. With the MOS-CuSum algorithm, the global detection delay is asymptotically (as the global ARLFA goes to infinity) minimized subject to constraints on both global ARLFA and average cost per sample at each sensor node, for any combination of the affected sensor nodes (Corollary 1).

Other Related Works: In the setting of quickest change detection with censoring sensors [21], [22], the sensors sample the observations with fixed information quality at each time instant and only send messages to the fusion center when the sampled observations satisfy certain conditions. The detection with control actions that can shape the information quality of observations taken (e.g., active sensing and sensor selection) has been extensively studied [3], [5], [23]–[25]. All of these works are in rather general settings: hypothesis testing or sequential hypothesis testing, while we focus on the quickest change detection, which is fundamentally different.

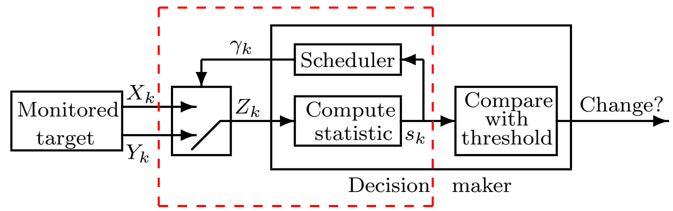


Fig. 1: Quickest change detection with observation scheduling.

Paper Organization: The remainder of this paper is organized as follows. The mathematical formulation of the considered problem is given in Section II. In Section III, we present the proposed algorithm and prove its asymptotic optimality. The observation scheduling in multi-channel setting is studied in Section IV, where an asymptotically optimal algorithm is given. Some numerical examples are provided in Section V and concluding remarks are given in Section VI.

Notations: $\mathbb{N}$, $\mathbb{N}_+$, $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ are the set of non-negative integers, positive integers, real numbers, non-negative real numbers and positive real numbers, respectively. $\mathbf{1}_A$ represents the indicator function that takes value 1 on the set A and 0 otherwise. $\times$ stands for the Cartesian product. For $x \in \mathbb{R}$, $(x)^+ = \max(0, x)$. Let X_r^n with $r, n \in \mathbb{N}_+$ denote the set $\{X_r, X_{r+1}, \ldots, X_n\}$. For $\vec{a}, \vec{b} \in \mathbb{R}^n$, let $\vec{a} \succeq (\succ) \vec{b}$ represent that $a_i \geq (>) b_i, 1 \leq i \leq n$.

II. Problem Setup

As illustrated in Fig. 1, there are two sequences of observations, $\{X_k\}_{k\in\mathbb{N}_+}$ and $\{Y_k\}_{k\in\mathbb{N}_+}$, about the monitored target. At each time instant k, the decision maker samples from either one. It is assumed that the observations before the change time ν, $X_1^{\nu-1}$ and $Y_1^{\nu-1}$, are independent and identically distributed (i.i.d.) with density f_X and f_Y, respectively. The observations after the change, X_ν^∞ and Y_ν^∞, are i.i.d. with density g_X and g_Y, respectively. It is further assumed that conditioned on the change time, $\{X_k\}$ and $\{Y_k\}$ are independent of each other for any $k \in \mathbb{N}_+$.

Let $\mathbf{I}(\cdot||\cdot)$ be the Kullback-Leibler (K-L) divergence [26], i.e., for probability density function (pdf) $f(x)$ and $g(x)$, $\mathbf{I}(f||g) := \int f(x) \ln \frac{f(x)}{g(x)} \mathrm{d}x$. Throughout this paper, we make the following assumption. The assumption is standard [20], [27] and avoids degenerate detection problem.

Assumption 1. *The following four quantities are finite and positive:* $\mathbf{I}(f_X||g_X), \mathbf{I}(g_X||f_X), \mathbf{I}(f_Y||g_Y)$ *and* $\mathbf{I}(g_Y||f_Y)$.

The schedule at time k is characterized by a binary variable γ_k as

$$\gamma_k = \begin{cases} 0, & \text{if } X_k \text{ is sampled}, \\ 1, & \text{if } Y_k \text{ is sampled}. \end{cases} \tag{1}$$

We assume that the decision maker must sample either X_k or Y_k, i.e., $\mathbf{1}_{\{\gamma_k=0\}} + \mathbf{1}_{\{\gamma_k=1\}} = 1$. Let Z_k denote the observation available at the decision maker at time k: $Z_k = X_k \mathbf{1}_{\{\gamma_k=0\}} + Y_k \mathbf{1}_{\{\gamma_k=1\}}$. Let θ denote the scheduling policy, i.e.,

$$\gamma_{k+1} = \theta(Z_1^k), k \in \mathbb{N}_+,$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

with γ_1 being determined by the *a priori* knowledge about the system. In addition to the stopping rule (the rule stating when the decision maker declares the change event), the scheduling policy θ should be determined.

Denote by $\mathbb{P}_\nu^\theta$ and $\mathbb{E}_\nu^\theta$ the probability measure and expectation when the scheduling policy θ is used and the change event happens at ν, and denote the same by $\mathbb{P}_\infty^\theta$ and $\mathbb{E}_\infty^\theta$ when there is no change. It is assumed that each sample X_k incurs cost c_X, while each sample Y_k incurs c_Y with $c_Y > c_X$. Let the cost of sampling at k be $c_Z = \mathbf{1}_{\{\gamma_k=0\}} c_X + \mathbf{1}_{\{\gamma_k=1\}} c_Y$. Let T be the stopping time, i.e., $\{T = k\}$ is measurable with respect to the σ-algebra generated by Z_1^k. The following average sampling cost constraint is considered:

$$c(\theta) = \limsup_{n\to\infty} \frac{1}{n} \mathbb{E}_\infty^\theta \left[\sum_{k=1}^{n} c_Z \middle| T > n \right] \leq \bar{c}, \tag{2}$$

where $c_X < \bar{c} \leq c_Y$ is a design parameter. The quantity $c(\theta)$ is the average cost per sample *before* the change event. Note that in the cost (2), which is inspired by the identical metric first proposed in [1] for data-efficient quickest change detection, we do not include the sampling cost for the interval between the actual change event and when it is detected. In practice, this additional cost usually can be neglected, since the detection delay is usually very small compared with the ARLFA. The asymptotic optimality result of the paper does not hold if the total cost is considered.

Given the pair of scheduling policy and stopping time (θ, T), there are two indices to characterize the detection performance: detection delay and false detection. For detection delay, we use Lorden's criterion [16]:

$$\begin{aligned} d_L(\theta, T) &= \sup_{1\leq\nu<\infty} d_L^\nu(\theta, T), \\ d_L^\nu(\theta, T) &:= \operatorname{ess\,sup} \mathbb{E}_\nu^\theta [(T-\nu+1)^+ | Z_1^{\nu-1}], \end{aligned} \tag{3}$$

with $Z_1^0 := \emptyset$ by convention. The delay $d_L(\theta, T)$ is largest delay one may encounter when the scheduling policy θ is used. It is considered in the worst-worst case (the least favorable realization of the observations and the change time). We should remark that the asymptotic optimality results (Theorem 2) also hold when Pollak's criterion [17] is used. It is well known that if $d_L(T, \Theta)$ is finite, false alarm is inevitable, i.e., $\mathbb{P}_\infty^\theta(T < \nu) = 1$ [16]. A reasonable measurement of false detection is the ARLFA: $\mathbb{E}_\infty^\theta[T]$. Note that $1/\mathbb{E}_\infty^\theta[T]$ can be interpreted as the frequency of false alarms.

We aim to find an optimal joint design of scheduling policy and stopping rule such that the detection delay is minimized subject to constraints on both false detection and average sensing cost. This problem is formally stated as follows.

Problem 1.

$$\begin{aligned} \underset{(\theta,T)}{\text{minimize}} \quad & d_L(\theta, T), \\ \text{subject to} \quad & \mathbb{E}_\infty^\theta[T] \geq \zeta, \qquad (4) \\ & c(\theta) \leq \bar{c}, \qquad (5) \end{aligned}$$

where $\zeta \geq 1$ is a given lower bound of the ARLFA and $c_X < \bar{c} \leq c_Y$ an upper bound on the average sampling cost.

III. Main Results

In this section, the proposed OS-CuSum algorithm is given and its properties are analyzed. The OS-CuSum algorithm is optimal in the asymptotic regime as the ARLFA goes to infinity. We should remark that it is too difficult to obtain the strictly optimal solution to Problem 1; see the relevant literatures [1], [12], [20]. In practice, scenarios where the probability of false alarm is sufficiently small (i.e., the ARLFA is sufficiently large) are more interesting.

A. The Observation Scheduling CuSum Algorithm

Let a, b be given parameters with $0 \leq b \leq a$. The stopping time and the observation scheduling policy are given by

$$T = \inf\{k : s_k \geq a\},$$
$$\theta(Z_1^k) = \begin{cases} 0, & \text{if } s_k < b, \\ 1, & \text{otherwise,} \end{cases}$$

where s_k evolves as

$$\begin{aligned} s_0 &= 0, \\ \tilde{s}_k &= (s_{k-1} + \ell(Z_k))^+, \\ s_k &= \begin{cases} b, & \text{if } \tilde{s}_k \geq b, s_{k-1} < b, \varepsilon_k \leq p, \\ 0, & \text{if } \tilde{s}_k \geq b, s_{k-1} < b, \varepsilon_k > p, \\ \tilde{s}_k, & \text{otherwise,} \end{cases} \end{aligned}$$

where $\ell(Z_k) = \ln \frac{g_X(Z_k)\mathbf{1}_{\{\gamma_k=0\}} + g_Y(Z_k)\mathbf{1}_{\{\gamma_k=1\}}}{f_X(Z_k)\mathbf{1}_{\{\gamma_k=0\}} + f_Y(Z_k)\mathbf{1}_{\{\gamma_k=1\}}}$ is the log-likelihood ratio of Z_k and $\varepsilon_k \sim \text{unif}(0, 1)$ is uniformly distributed and independent of X_k or Y_k.

Remark 1. *The OS-CuSum algorithm is a variant of the CuSum algorithm with stochastically switching observation modes. If $b = 0$ ($b = a$), the algorithm reduces to the CuSum algorithm with $Z_k = Y_k$ ($Z_k = X_k$), $\forall k \in \mathbb{N}_+$. The random variable ε_k is introduced to make s_k cross b stochastically, which makes it possible for the OS-CuSum algorithm to satisfy any average sampling cost constraint as well as to attain the asymptotic optimality (see proof of Theorem 2). If $p = 1$, s_k crosses b deterministically and the sampling cost cannot be set arbitrarily by only adjusting b.*

Remark 2. *Here we compare our algorithm with the DE-CuSum algorithm proposed in [1]. The DE-CuSum algorithm is a variant of the CuSum algorithm that allows the detection statistic to be negative. When the statistic is negative, it does not update with the likelihood ratio of the observations (and the sensor does not sample) but increases with a constant each time instant. The DE-CuSum algorithm is for the "data-efficient" scenario and one cannot apply it in our case. This is because when the statistic increases with a constant, it does not utilize the information of the observations at all, which will inevitably cause performance deterioration when the ARLFA takes moderate values. Efficiently utilizing the information of both sequences of observations to satisfy the sampling cost constraint as well as to possess good detection performance is challenging for the algorithm design. We thus introduce the stochastic crossing mechanism. We should remark that such stochastic crossing mechanism has not been found in the quickest change detection related literatures. Note that*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

not sampling can be viewed as a special observation, of which both the K–L divergence and cost are zero. Our algorithm can be slightly modified to work in the "data-efficient" scenario. One may modify it as follows. Let $b = 0$ and the number of the time instants s_k stays at zero, once it becomes zero, be geometrically distributed with parameter p that can be adjusted. Note that the undershoot in the statistics is utilized in the DE-CuSum algorithm, while the OS-CuSum algorithm does not use such information. Thus compared to the DE-CuSum algorithm, the OS-CuSum algorithm (specialized to data-efficient scenarios) might suffer from some performance loss.

Remark 3. *It might happen that our algorithm cannot satisfy the average cost per sample constraint when $\bar{c}$ is close to c_Y. This is because for any $b > 0$, $\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_\infty^\theta\left[\sum_{k=1}^n \mathbf{1}_{\{\gamma_k=0\}}\middle| T > n\right] > \lim_{k\to\infty}\mathbb{P}_\nu^\theta\{s_k = 0\} > 0$. To cope with this, one can adopt "stochastic restart", i.e., $\gamma_{k+1} = 0$ with certain probability (less than one) whenever $s_k = 0$. The main results (Theorems 1–3 and Corollary 1) can be easily generalized to the case with stochastic restart. In practice, however, the more interesting situation is when the average cost per sample constraint is severe, i.e., $\bar{c}$ is close to c_X. We only study the algorithm with deterministic restart for simplicity of presentation.*

In the remainder of this paper, the stopping time of the OS-CuSum algorithm is referred to as $T_{(a,b,p)}$. Since the algorithm is completely characterized by (a, b, p), to highlight the specific values of these parameters, we also use $T_{(a,b,p)}$ to represent the algorithm. The properties of the OS-CuSum algorithm are given in the following.

B. Asymptotic Optimality

Theorem 1. *The OS-CuSum algorithm is an equalizer rule, i.e., $d_L(T_{(a,b,p)}) = d_L^\nu(T_{(a,b,p)}), \forall \nu \in \mathbb{N}_+$.*

Proof: When $b = 0$ or $b = a$, the OS-CuSum algorithm reduces to the CuSum algorithm. It is well known that the CuSum algorithm is an equalizer rule [10].

When $0 < b < a$, following the same reasoning as in proof of Lemma 1 and Theorem 1 in [22], one obtains the desired result. ∎

Remark 4. *This property is beneficial to reduce the computational burden to determine the parameters of the algorithm. In general, the parameters can only be determined by numerically evaluating the performance indices, which include $d_L(\theta, T)$. The property of Theorem 1, however, means that for simplicity, one can just let $\nu = 1$.*

Before presenting the main theorem, we first give a supporting lemma that provides the asymptotic lower bound of detection delay for any scheduling policy and detection procedure. The proof is presented in Appendix A.

Lemma 1. *For any $c_X < \bar{c} \le c_Y$, scheduling policy θ and stopping time T, as $\zeta \to \infty$,*

$$\inf\{d_L(\theta, T) : \mathbb{E}_\infty^\theta[T] \ge \zeta\} \ge \frac{\ln \zeta}{\max(\mathbf{I}(g_X||f_X), \mathbf{I}(g_Y||f_Y))}(1 + o(1)). \quad (6)$$

Note that since $c_X < \bar{c} \le c_Y$, if $\mathbf{I}(g_X||f_X) \ge \mathbf{I}(g_Y||f_Y)$, to achieve the asymptotic optimality, the observation X_k can be sampled along the whole horizon, i.e., $T_{(\ln\zeta,\ln\zeta,p)}$ is asymptotically optimal. We henceforth assume that $\mathbf{I}(g_X||f_X) < \mathbf{I}(g_Y||f_Y)$.

Theorem 2. *For any $c_X < \bar{c} \le c_Y$, $\zeta > 1$ and $0 < b < \ln\zeta$, there exists $0 < p^* \le 1$ such that $T_{(\ln\zeta,b,p)}, p \in (0, p^*]$, satisfies the constraints (4) and (5). What is more, $T_{(\ln\zeta,b,p)}$ with $p \in (0, p^*]$ is asymptotically optimal, i.e., as $\zeta \to \infty$*

$$d_L(T_{(\ln\zeta,b,p)}) \le \frac{\ln\zeta}{\mathbf{I}(g_Y||f_Y)}(1 + o(1)). \quad (7)$$

Proof: See Appendix B. ∎

Remark 5. *The mechanism of stochastically crossing b from below (when $p < 1$) is introduced to satisfy any average sampling cost constraints as well as to attain the asymptotic optimality. This stochastic crossing mechanism increases (compared with $p = 1$) the duration when $s_k \in [0, b]$, which, alas, inevitably increases the detection delay $\mathbb{E}_1^\theta[T_{(\ln\zeta,b,p)}]$. Fortunately, given p, the increment of $\mathbb{E}_\infty^\theta[T_{(\ln\zeta,b,p)}]$ generally significantly exceeds that of $\mathbb{E}_1^\theta[T_{(\ln\zeta,b,p)}]$, which means that the ARLFA constraint (4) can be satisfied with a relatively small cost (the detection delay is increased). This can be verified using Wald's approximation (Page 11, [28]), which is an approximation method that relates the expected sample size with the two thresholds for a two-sided sequential probability ratio test. By Wald's approximation, $\mathbb{E}_\infty^\theta[T_{(b,b,p)}] = \frac{|e^b - b - 1|}{p\mathbf{I}(f_X||g_X)}$ and $\mathbb{E}_1^\theta[T_{(b,b,p)}] = \frac{|e^{-b} + b - 1|}{p\mathbf{I}(g_X||f_X)}$.*

Remark 6. *One can see in the proof that the OS-CuSum algorithm can be easily generalized to the scenario where multiple sequences of observations are available. The OS-CuSum algorithm might be modified as follows: the sequence of observations that has the largest K-L divergence, say $\mathbf{I}_{\max}(g||f)$, are sampled when $s_k \ge b$, and any other observations are sampled when $s_k < b$. It can be verified that the asymptotically optimal detection delay only depends on $\mathbf{I}_{\max}(g||f)$.*

IV. Multi-channel Observation Scheduling

In this section, we study the quickest change detection with observation scheduling in the multi-channel setting, which is illustrated in Fig. 2. Suppose that the target is monitored by M sensors. Denote by $\mathcal{M} = \{1, 2, \dots, M\}$ the set of sensors. Sensor $m \in \mathcal{M}$ can access $X_{m,k}$ or $Y_{m,k}$ and sends a message to the fusion center at each time instant. At an unknown time instant ν, an event occurs and the distribution of the observations for an unknown subset of the sensor nodes change. Let $\Xi \subseteq \mathcal{M}$ be the set of sensors that are affected. It is assumed that at sensor $m \in \Xi$, the observations before the change $X_{m,1}^{\nu-1}$ ($Y_{m,1}^{\nu-1}$) are i.i.d. with density $f_{X,m}$ ($f_{Y,m}$) and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

the observations after the change $X_{m,\nu}^{\infty}$ ($Y_{m,\nu}^{\infty}$) are i.i.d. with $g_{X,m}$ ($g_{Y,m}$). For a sensor $m \notin \Xi$, the observations $\{X_{m,k}\}$ ($\{Y_{m,k}\}$) are i.i.d. with density $f_{X,m}$ ($f_{Y,m}$) along the whole horizon. It is assumed that the observations are independent across the sensors. We aim to find the optimal stopping time at the fusion center and the optimal observation scheduling policy for each sensor node.

In addition to the finiteness and positiveness assumption of the K–L divergence of pre-change and post-change distributions at each sensor node as in Assumption 1, for the multi-channel scenario we further assume that the second moment of the K–L divergence of $\{Y_{m,k}\}$ is finite:

$$\int g_{Y,m}(x)\left(\ln\frac{g_{Y,m}(x)}{f_{Y,m}(x)}\right)^2 \mathrm{d}x < \infty. \tag{8}$$

This is a technical requirement for Theorem 3 to bound the asymptotic detection delay. Similar assumptions can be found in [2], [13], [27].

Let $\gamma_{m,k}, Z_{m,k}, \theta_m, c_{m,X}, c_{m,Y}, c_{m,Z}$ be the corresponding parts of $\gamma_k, Z_k, \theta, c_X, c_Y, c_Z$ defined in Section II for sensor m. Let $\Theta = \{\theta_1, \dots, \theta_M\}$ be the observation scheduling policy for the whole sensor network. Denote by $\mathbb{P}_{\nu}^{\Theta,\Xi}$ and $\mathbb{E}_{\nu}^{\Theta,\Xi}$ the probability measure and expectation when the scheduling policy Θ is used, the change event happens at ν and the sensor $m \in \Xi$ is affected, and denote the same by $\mathbb{P}_{\infty}^{\Theta,\Xi}$ and $\mathbb{E}_{\infty}^{\Theta,\Xi}$ when there is no change. Let T be the stopping time at the fusion center. We consider the constraint on average sampling cost before the change for each sensor:

$$c_m(\Theta) = \limsup_{n\to\infty}\frac{1}{n}\mathbb{E}_{\infty}^{\Theta,\Xi}\left[\sum_{k=1}^{n} c_{m,Z}\middle| T > n\right] \le \bar{c}_m, \tag{9}$$

where $c_{m,X} < \bar{c}_m \le c_{m,Y}$ is a design parameter. Let $\mathbf{Z}_k = \{Z_{1,k}, \dots, Z_{M,k}\}$. The Lorden's detection delay is given by

$$d_L(\Theta, T) = \sup_{1\le\nu<\infty} \operatorname{ess\,sup} \mathbb{E}_{\nu}^{\Theta,\Xi}[(T-\nu+1)^+ | \mathbf{Z}_1^{\nu-1}].$$

The problem of optimal joint design of multi-channel observation scheduling policy and stopping rule is stated as follows:

Problem 2. *For any non-empty $\Xi \subseteq \mathcal{M}$,*

$$\begin{aligned} \underset{(\Theta,T)}{\text{minimize}} \quad & d_L(\Theta, T), \\ \text{subject to} \quad & \mathbb{E}_{\infty}^{\Theta,\Xi}[T] \ge \zeta, && (10)\\ & c_m(\Theta) \le \bar{c}_m, \forall m \in \mathcal{M}, && (11) \end{aligned}$$

where $\zeta \ge 1$ is a given lower bound of the global ARLFA and $c_{m,X} < \bar{c}_m \le c_{m,Y}$ is an upper bound on average sampling cost for the sensor m.

The MOS-CuSum algorithm is as follows. Each sensor runs the observation scheduling policy proposed in Section III locally. The stopping rule N_{sum} proposed in [20] is used at the fusion center. The details are as follows. The observation scheduling policy at sensor m is given by

$$\theta_m(Z_{m,1}^k) = \begin{cases} 0, & \text{if } s_{m,k} < b_m , \\ 1, & \text{if } s_{m,k} \ge b_m , \end{cases} \tag{12}$$

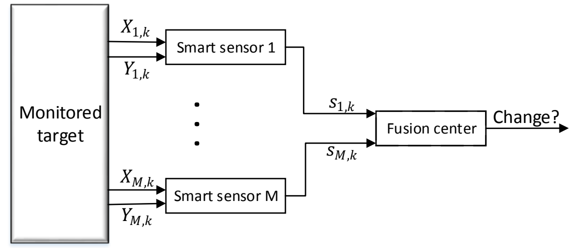


Fig. 2: Quickest change detection with observation scheduling in a multi-channel setting. Each smart sensor corresponds to the red-dashed rectangle in Fig. 1.

where $s_{m,k}$ evolves as

$$\begin{aligned} s_{m,0} &= 0, \\ \tilde{s}_{m,k} &= \left(s_{m,k-1} + \ell(Z_{m,k})\right)^+, \\ s_{m,k} &= \begin{cases} b_m, & \text{if } \tilde{s}_{m,k} \ge b_m, s_{m,k-1} < b_m, \varepsilon_{m,k} \le p_m, \\ 0, & \text{if } \tilde{s}_{m,k} \ge b_m, s_{m,k-1} < b_m, \varepsilon_{m,k} > p_m, \\ \tilde{s}_{m,k}, & \text{otherwise,} \end{cases} \end{aligned}$$

with $\varepsilon_{m,k} \sim \text{unif}(0,1)$ being uniformly distributed and independent of $X_{m,k}$ or $Y_{m,k}$. The stopping time at the fusion center is

$$T = \inf\{k : \sum_{m=1}^{M} s_{m,k} \ge a\}.$$

Remark 7. *Because the observation scheduling policy runs locally at each sensor node, the algorithm inherits the scalability of the N_{sum} procedure, see [20].*

In the following, we show the asymptotic optimality of the MOS-CuSum algorithm for any possible non-empty Ξ, which is a generalization of the results obtained in [20]. With a little abuse of notation, let $b = [b_1, \dots, b_m]$ and $p = [p_1, \dots, p_m]$. To specifically point out the parameters used, as for the OS-CuSum algorithm, we henceforth call the MOS-CuSum algorithm (also the stopping time) $T(a, b, p)$. As in Section III, we assume that $\mathbf{I}(g_{Y,m}||f_{Y,m}) \ge \mathbf{I}(g_{X,m}||f_{X,m}), \forall m \in \mathcal{M}$.

Theorem 3. *For any possible non-empty $\Xi \subseteq \mathcal{M}$, finite $b \succeq 0$ and $0 \prec p \preceq 1$, as $a \to \infty$,*

$$d_L(T(a,b,p)) \le \frac{a}{\sum_{m\in\Xi}\mathbf{I}(g_{Y,m}||f_{Y,m})}(1+o(1)), \tag{13}$$

$$\mathbb{E}_{\infty}^{\Theta,\Xi}[T(a,b,p)] \ge \frac{e^a}{1+a+a^2/2!+\cdots+a^{M-1}/(M-1)!}. \tag{14}$$

Proof: See Appendix C. ■

Remark 8. *In [2], a similar algorithm called DE-Censor-Sum algorithm is studied. In the DE-Censor-Sum algorithm, the DE-CuSum algorithm proposed in [1] runs locally at the sensor nodes and the N_{sum} procedure in [20] is used at the fusion center. Despite the similarities, the methodology used to bound the asymptotic detection delay of the MOS-CuSum algorithm (i.e., equation* (13)*) is completely different from that in [2]. We remark that our method is much simpler, and since our method also works in the scenarios studied in [2], it provides another perspective on the asymptotic detection delay of*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

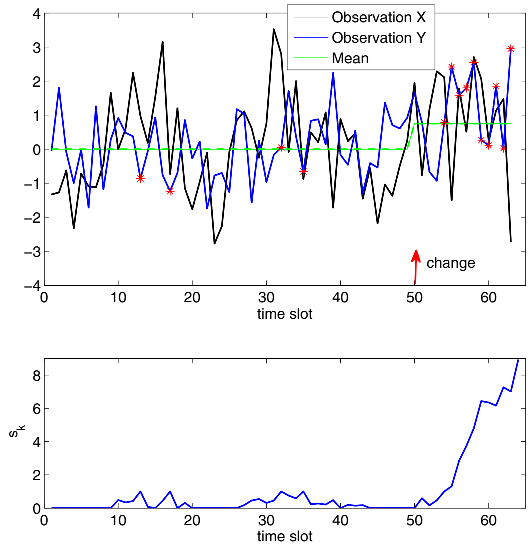


Fig. 3: Typical evolution of the OS-CuSum algorithm. In the top plot, the red stars are used to indicate when the observation Y_k is sampled.

the DE-Censor-Sum algorithm. Also, compared with the DE-Censor-Sum algorithm, our algorithm requires slightly more effort to bound its asymptotic ARLFA (i.e., equation (14)*).*

Corollary 1. *Let* $a_\zeta = \ln\zeta + (M-1)\ln\ln\zeta$. *Then for any non-empty* $\Xi \subseteq \mathcal{M}$, $c_{m,X} < \bar{c}_m \le c_{m,Y}$ *and finite* b, *there exists* $0 \prec p^* \preceq 1$ *such that for any* $p \preceq p^*$, *as* $\zeta \to \infty$, $T(a_\zeta, b, p)$ *satisfies* (10) *and sampling cost constraints* (11) *for any* $\bar{c}_m$, *and also minimizes the detection delay, i.e.,*

$$d_L(T(a_\zeta, b, p)) \le \frac{\ln\zeta}{\sum_{m\in\Xi}\mathbf{I}(g_{Y,m}||f_{Y,m})}(1+o(1)).$$

Proof: The existence of p^* such that the sampling cost constraint is satisfied can be easily verified using the similar analysis of the proof of Lemma 5.

By (14), as in Corollary 1 of [20], it is straightforward to show that as $\zeta \to \infty$, $T(a_\zeta, b, p)$ satisfies (10).

One can easily generalize Lemma 1 to the decentralized case. That is, given any non-empty Ξ and $\bar{c}_m$, for any (Θ, T), as $\zeta \to \infty$, if $\mathbb{E}_\infty^{\Theta,\Xi}[T] \ge \zeta$, then

$$d_L(\Theta, T) \ge \frac{\ln\zeta}{\sum_{m\in\Xi}\mathbf{I}(g_{Y,m}||f_{Y,m})}(1+o(1)).$$

The asymptotic optimality of $T(a_\zeta, b, p)$ thus follows directly from (13). ■

V. Numerical Examples

In this section, first we illustrate the behavior of the proposed OS-CuSum and MOS-CuSum algorithms through a numerical example. Then we use two examples to illustrate the asymptotic optimality of the OS-CuSum (Theorem 2) and the MOS-CuSum (Corollary 1). Lastly, we compare the OS-CuSum algorithm with two simple heuristic schemes: one periodic scheme and one stochastic scheme.

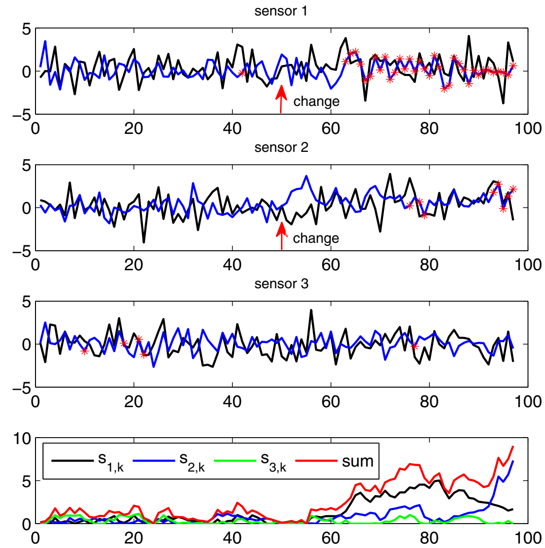


Fig. 4: Typical evolution of the MOS-CuSum algorithm. The top three plots are plotted as in Fig. 3: the black curve represents the observation $X_{m,k}$, the blue one represents the observation $Y_{m,k}$ and the red stars are used to indicate when the observation $Y_{m,k}$ is sampled.

Example 1. To illustrate the OS-CuSum algorithm, it is assumed that $f_X \sim \mathcal{N}(0,2), g_X \sim \mathcal{N}(0.75,2), f_Y \sim \mathcal{N}(0,1)$ and $g_Y \sim \mathcal{N}(0.75,1)$. We assume that the change event happens at $k=50$ and let $a=8, b=1, p=0.95$. For the MOS-CuSum algorithm, it is assumed that $M=3$ sensors are deployed: for $m \in \{1,2,3\}$, $f_{X,m} \sim \mathcal{N}(0,2), g_{X,m} \sim \mathcal{N}(0.5,2)$, for $m \in \{1,3\}$, $f_{Y,m} \sim \mathcal{N}(0,1), g_{Y,m} \sim \mathcal{N}(0.5,1)$ and $f_{Y,2} \sim \mathcal{N}(0,1), g_{Y,2} \sim \mathcal{N}(1,1)$. It is further assumed that when the change event happens, only sensors 1 and 2 are affected. Let $b=[1,2,1]$, $p=[1,1,1]$ and $a=8$. As illustrated in Fig. 3 and Fig. 4, the underlying mechanism of the observation scheduling policy can be heuristically stated as follows: before the change event happens, at most times the inexpensive observations X_k (or $X_{m,k}$) are sampled to satisfy the sampling cost constraint, while after the change event happens (this period is always quite short, since it is just the detection delay we aim to minimize), the expensive but informative observations Y_k (or $Y_{m,k}$) are sampled at most times to minimize the detection delay. In Fig. 4, one can also see that compared with sensor 1, sensor 2 samples less informative observations ($Y_{2,k}$) due to that b_2 is larger than b_1. The detection statistic $s_{2,k}$, however, eventually dominates $s_{1,k}$ and $s_{3,k}$ as illustrated in the bottom plot. Heuristically, it is because $Y_{2,k}$ contains more information than $Y_{1,k}$, i.e., $\mathbf{I}(g_{Y,2}||f_{Y,2}) > \mathbf{I}(g_{Y,1}||f_{Y,1})$. We conclude that this example illustrates how the OS-CuSum and the MOS-CuSum algorithms are able to automatically schedule the sampling attention to the appropriate observations and sensors.

Example 2. We use the same f_X, g_X, f_Y and g_Y as in Example 1. By simple calculation, one obtain that $2\mathbf{I}(g_X||f_X) = \mathbf{I}(g_Y||f_Y) = 9/32$. The other parameters are as follows: $c_X = 1, c_Y = 1.5, \bar{c} = 1.1$. Given constraints on both the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

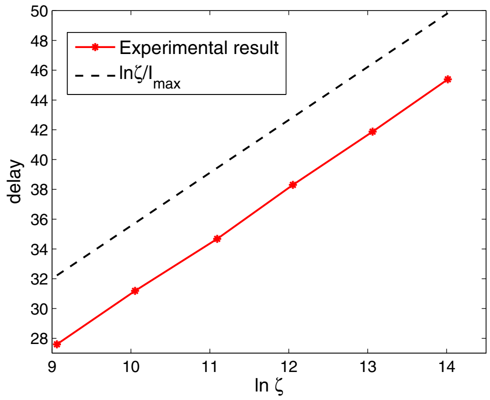


Fig. 5: Detection delay versus the ARLFA for the OS-CuSum algorithm.

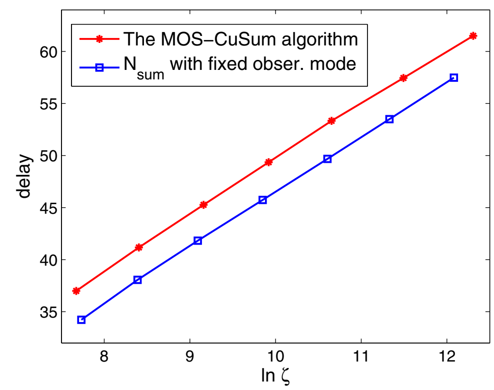


Fig. 6: Detection delay versus the ARLFA for the MOS-CuSum algorithm.

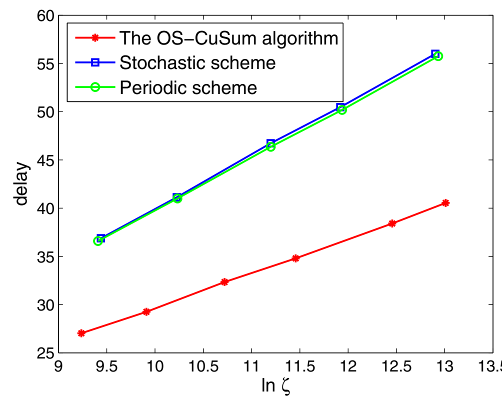


Fig. 7: Detection delay versus the ARLFA for the proposed OS-CuSum algorithm and two heuristic algorithms.

ARLFA and average cost per sample, there exist multiple admissible parameter pairs (a, b, p) for the proposed algorithm. Fix $b = 0.68, p = 0.95$ and vary a (sufficiently large) to obtain different ARLFA and delays (as long as a is sufficiently large, the average cost per sample remains almost the same). As shown in Fig. 5, the experimental curve and the curve for theoretically asymptotic lower bound of detection delay are parallel, which indicates the asymptotic optimality of the OS-CuSum algorithm (Theorem 2), since as $\zeta \to \infty$, the delay gap is negligible. Note that with the same ARLFA, our algorithm has smaller detection delay than the theoretic lower bound. This is acceptable because $\ln \zeta/\mathbf{I}_{\max}$ can is a lower bound of detection delay only in the asymptotic regime. One can find a similar simulation result in Fig. 3 of [12].

Example 3. In this example, the asymptotic optimality of the MOS-CuSuam algorithm (Corollary 1) is illustrated. It is assumed that $M = 5$ identical sensors are deployed, and the same $f_{X,m}, g_{X,m}, f_{Y,m}$ and $g_{Y,m}$ as for sensor 1 and 3 in Example 1 are used. It is assumed that sensor 1 and sensor 2 are affected when the change event happens, but the fusion center is not aware of this. We let $p_m = 1, \forall m$. Note that the value of p_m only affects the average sampling cost and does not affect the asymptotic optimality of the MOS-CuSum algorithm. As in Example 2, we fix b_m for each sensor and vary a to obtain the curve for our algorithm, which is shown in Fig. 6. We use the N_{sum} algorithm with fixed observation mode (the observation $\{Y_{m,k}\}$ is used at each time instant for each sensor) for comparison. Since the N_{sum} algorithm is asymptotically optimal [20] and the observation $\{Y_{m,k}\}$ is more favorable than $\{X_{m,k}\}$, the detection performance of the N_{sum} algorithm with observation $\{Y_{m,k}\}$ is the limit of our algorithm. Fig. 6 verifies the asymptotic optimality of our algorithm since the two curves are parallel.

Example 4. In this example, we illustrate the detection performance of the OS-CuSum algorithm by comparing it with a periodic scheme and a stochastic scheme. In the both schemes, the conventional CuSum algorithm is used as the detection procedure and the observation scheduling policies are independent of the realization of observations. Note that the energy constraint in (2) can be reinterpreted as a bound of the ratio the number of Y_k to that of X_k. Suppose that $\limsup_{n\to\infty} \mathbb{E}_\infty^\theta[\sum_{k=1}^n \mathbf{1}_{\{\gamma_k=1\}}]/\mathbb{E}_\infty^\theta[\sum_{k=1}^n \mathbf{1}_{\{\gamma_k=0\}}] \le q_1/q_2$ (assume that q_1/q_2 is a simple fraction), then the observation scheduling policy for the periodic scheme is give by

$$\gamma_k = \begin{cases} 0, & \text{if } k \bmod (q_1+q_2) < q_2, \\ 1, & \text{otherwise}, \end{cases} \tag{15}$$

and the observation scheduling policy for the stochastic scheme is as follows:

$$\gamma_k = \begin{cases} 0, & \text{if } \tilde{\varepsilon}_k \le q_2/(q_1+q_2), \\ 1, & \text{otherwise}, \end{cases} \tag{16}$$

where $\tilde{\varepsilon}_k \sim \text{unif}(0,1)$ is uniformly distributed and independent of the observations. Let f_X, g_X, f_Y and g_Y be the same as in Example 1. Let $q_1/q_2 = 3/7$. Fig. 7 shows that the OS-CuSum algorithm significantly outperforms the periodic scheme and the stochastic one. It should be noted that as the ARLFA increases, the delay difference between our algorithm and the other two algorithms increases. This is because the OS-CuSum algorithm is asymptotically optimal while the other two are not. The detection difference goes to infinity as the ARLFA goes to infinity.

VI. Conclusion and Future Works

In this paper, we have studied the quickest change detection with observation scheduling in the minimax setting. Observation sequences with different cost and information quality

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

are available at the decision maker, but the decision maker can access only one of them due to various limitations. We considered the Lorden's criterion for the detection delay. An algorithm was proposed: the observation scheduling policy has a threshold structure and the detection procedure is a variant of the CuSum algorithm where the detection statistic stochastically crosses the threshold used for the scheduling policy. We proved that the algorithm asymptotically minimizes (as the ARLFA goes to infinity) the detection delay for any average cost per sample constraint. We further studied the quickest change detection with observation scheduling in the multi-channel setting. We show that when each sensor uses the proposed observation scheduling policy locally and the fusion center uses the N_{sum} algorithm, the global detection delay is asymptotically minimized for any average cost per sample constraint for each sensor node and any possible combination of the sensors that are affected by the change event.

Future work includes studying the problem in the Bayesian setting and considering total cost of sampling constraint instead of average cost.

Appendix A
Proof of Lemma 1

Let $\mathbf{I}_{\max} = \max(\mathbf{I}(g_X||f_X), \mathbf{I}(g_Y||f_Y))$. Before proceeding, we firs give the following two supporting lemmas. In Lemma 2, we show that for any possible scheduling policy and realizations, the long averaged log likelihood ratio of the available observations is bounded by the term $\mathbf{I}_{\max}$.

Lemma 2. *For any $\nu \geq 1$, realization $Z_1^{\nu-1}$ and scheduling policy θ, as $n \to \infty$*

$$\frac{1}{n}\sum_{k=\nu}^{\nu+n-1} \ell(Z_k) \leq \mathbf{I}_{\max}$$

holds almost surely under $\mathbb{P}_\nu^\theta$.

Proof: Define a sequence of random variables $\{k_X(i)\}_{i\in\mathbb{N}_+}$ as

$$k_X(1) = \inf\{k : k \geq \nu, \gamma_k = 0\},$$
$$k_X(i) = \inf\{k : k > k_X(i-1), \gamma_k = 0\}, \forall i \geq 2,$$

and let $\bar{i}_X(n) = \max\{i : k_X(i) < n\}$. Note that the term $\bar{i}_X(n)$ is the number of observations $\{X_k\}$ are sampled between time instants ν and n. The quantities $\{k_Y(i)\}_{i\in\mathbb{N}_+}$ and $\bar{i}_Y(n)$ are defined similarly. Then one can see that for any $Z_1^{\nu-1}$ and θ, $Z_{k_X(i)}$ ($Z_{k_Y(i)}$) are i.i.d. ($\{Z_k\}$ of course is not i.i.d., but the dependence of $\{Z_k\}$ is captured by the values of $\{k_X(i)\}$ and $\{k_Y(i)\}$) with density g_X (g_Y) under $\mathbb{P}_\nu^\theta$. The strong law of large number (SLLN) yields that under $\mathbb{P}_\nu^\theta$,

$$\lim_{\bar{i}_X(n)\to\infty} \frac{\ell(Z_{k_X(1)}) + \cdots + \ell(Z_{k_X(\bar{i}_X(n))})}{\bar{i}_X(n)} \overset{a.s.}{\to} \mathbf{I}(g_X||f_X),$$
$$\lim_{\bar{i}_Y(n)\to\infty} \frac{\ell(Z_{k_Y(1)}) + \cdots + \ell(Z_{k_Y(\bar{i}_Y(n))})}{\bar{i}_Y(n)} \overset{a.s.}{\to} \mathbf{I}(g_Y||f_Y).$$

Let $\alpha_X = \lim_{n\to\infty}\frac{\bar{i}_X(n)}{n}$ and $\alpha_Y = \lim_{n\to\infty}\frac{\bar{i}_Y(n)}{n}$. Then by Assumption 1, under $\mathbb{P}_\nu^\theta$, as $n \to \infty$

$$\begin{aligned}&\frac{1}{n}\sum_{k=\nu}^{\nu+n-1}\ell(Z_k)\\ &= \frac{\bar{i}_X(n)}{n}\frac{1}{\bar{i}_X(n)}\sum_{i=1}^{\bar{i}_X(n)} Z_{k_X(i)} + \frac{\bar{i}_Y(n)}{n}\frac{1}{\bar{i}_Y(n)}\sum_{i=1}^{\bar{i}_Y(n)} Z_{k_Y(i)}\\ &\overset{a.s.}{\to} \begin{cases} \mathbf{I}(g_X||f_X), & \text{if } i_Y^\infty = 0,\\ \mathbf{I}(g_Y||f_Y), & \text{if } i_X^\infty = 0,\\ \alpha_X\mathbf{I}(g_X||f_X) + \alpha_Y\mathbf{I}(g_Y||f_Y), & \text{otherwise.}\end{cases}\end{aligned}$$

Note that $\alpha_X + \alpha_Y = 1$, the desired result follows easily. ■

For any $\nu \geq 1$ and given n, define

$$t_\nu^n = \arg\max_{t\leq n}\sum_{k=\nu}^{\nu+t}\ell(Z_k)$$

as the time that the random walk (with each step taking $\ell(Z_k)$) reaches its maximum between ν and n. Then about the asymptotic behavior of t_ν^n, we have the following lemma.

Lemma 3. *For any $\nu \geq 1$, realization $Z_1^{\nu-1}$ and scheduling policy θ, as $n \to \infty$*

$$t_\nu^n \to \infty$$

holds with probability one under $\mathbb{P}_\nu^\theta$.

Proof: Let $t_\Delta = n - t_\nu^n$. Then by definition of t_ν^n,

$$\mathbb{P}_\nu^\theta\{t_\Delta = i\} < \mathbb{P}_\nu^\theta\left\{\sum_{k=\nu+t_\nu^n+1}^{\nu+t_\nu^n+i}\ell(Z_k) < 0\right\}.$$

By Lemma 2, for any ν, t_ν^n and θ, as $t_\Delta \to \infty$,

$$\sum_{k=\nu+t_\nu^n+1}^{\nu+t_\nu^n+t_\Delta}\ell(Z_k) \overset{a.s.}{\to} \infty,$$

which implies that $\mathbb{P}_\nu^\theta\{t_\Delta < \infty\} = 1$. The desired result thus follows. ■

We are ready to prove Lemma 1. By theorem 1 of [29], to obtain (6), it suffices to prove that given an arbitrary scheduling policy θ, for any $\delta > 0$

$$\begin{aligned}&\lim_{n\to\infty}\sup_{\nu\geq 1}\operatorname{ess\,sup}\mathbb{P}_\nu^\theta\Big\{\max_{t\leq n}\sum_{k=\nu}^{\nu+t}\ell(Z_k) \geq \mathbf{I}_{\max}(1+\delta)n\big|Z_1^{\nu-1}\Big\}\\ &= 0.\end{aligned} \qquad (17)$$

For any $\nu \geq 1$,

$$\begin{aligned}
&\lim_{n\to\infty} \operatorname{ess\,sup} \mathbb{P}_\nu^\theta\Big\{ \max_{t\leq n} \sum_{k=\nu}^{\nu+t} \ell(Z_k) \geq \mathbf{I}_{\max}(1+\delta)n \big| Z_1^{\nu-1}\Big\}\\
&= \lim_{n\to\infty} \operatorname{ess\,sup} \mathbb{P}_\nu^\theta\Big\{ \frac{1}{n} \sum_{k=\nu}^{\nu+t_\nu^n} \ell(Z_k) \geq \mathbf{I}_{\max}(1+\delta) \big| Z_1^{\nu-1}\Big\}\\
&\leq \lim_{n\to\infty} \operatorname{ess\,sup} \mathbb{P}_\nu^\theta\Big\{ \frac{1}{t_\nu^n} \sum_{k=\nu}^{\nu+t_\nu^n} \ell(Z_k) \geq \mathbf{I}_{\max}(1+\delta) \big| Z_1^{\nu-1}\Big\}\\
&= \lim_{t_\nu^n\to\infty} \operatorname{ess\,sup} \mathbb{P}_\nu^\theta\Big\{ \frac{1}{t_\nu^n} \sum_{k=\nu}^{\nu+t_\nu^n} \ell(Z_k) \geq \mathbf{I}_{\max}(1+\delta) \big| Z_1^{\nu-1}\Big\}\\
&= 0,
\end{aligned}$$

where the second equality follows from Lemma 3 and the last equality follows from Lemma 2. The relation (17) thus follows.

Appendix B
Proof of Theorem 2

This proof is done by relating our algorithm $T_{(\ln\zeta,b,p)}$ to the classical CuSum algorithm $T_{(\ln\zeta,0,p)}$. The algorithm $T_{(\ln\zeta,b,p)}$ is treated as a sequence of two-sided sequential probability ratio tests (SPRTs) with two different distributions (since the observation modes are switching).

We first show that for any switching threshold b, by appropriately choosing the successful crossing probability p, the ARLFA of our algorithm can be larger than that of the classical CuSum algorithm with same and arbitrary stopping threshold a. What is more, p can be chosen independent of a.

Lemma 4. *For any $0 < b < a$, there exists $0 < \hat{p} < 1$ such that $\hat{p}$ only depends on b and for any $p \in (0, \hat{p}]$, $\mathbb{E}_\infty^\theta[T_{(a,b,p)}] \geq \mathbb{E}_\infty^\theta[T_{(a,0,p)}]$.*

Proof: Let $T^c_{(a,b,\omega)}$ be the stopping time $T^c_{(a,b,\omega)} = \inf\{k : s_k^c \geq a\}$, where s_k^c starts at $\omega \geq 0$ and evolves as

$$\tilde{s}_k^c = \left(s_{k-1}^c + \ell(Y_k)\right)^+,$$
$$s_k^c = \begin{cases} b, & \text{if } \tilde{s}_k^c \geq b, s_{k-1}^c < b,\\ \tilde{s}_k^c, & \text{otherwise.}\end{cases}$$

Note that here the observations $\{Y_k\}$ are used at each time instant. Let s_k^* be the detection statistic of the algorithm $T_{(a,0,p)}$. Note that due to the reset action when crossing b from below and the nonegative initial value ω for s_k^c, for any realization Y_1^k and $k \geq 1$, $s_k^c \leq s_k^*$, the following thus holds:

$$\mathbb{E}_\infty^\theta[T^c_{(a,b,0)}] \geq \mathbb{E}_\infty[T_{(a,0,p)}]. \tag{18}$$

Define

$$\phi = \inf\{i : \sum_{k=1}^{i} \ell(Z_k) < 0 \text{ or} \sum_{k=1}^{i} \ell(Z_k) > a - b\}, \tag{19}$$

$$\psi = \left(b + \sum_{k=1}^{\phi} \ell(Z_k)\right)^+. \tag{20}$$

Note that ϕ is a stopping time indicating how long s_k stays in $[a, b]$ each time when it enters into this interval, and ψ is the value s_k takes when it jumps out. Let $W_{(a,b,p,\omega)}$ be a stopping time defined in the same manner as $T_{(a,b,p)}$ but with the initial start $s_0 = \omega \geq 0$.

Interpreting both $T_{(a,b,p)}$ and $T^c_{(a,b,0)}$ as a sequence of two-sided SPRTs and by Wald's identity (Page 12, [28]), one obtains that

$$\begin{aligned}
&\mathbb{E}_\infty^\theta[T^c_{(a,b,0)}]\\
&= \mathbb{E}_\infty^\theta[T^c_{(b,0,0)}] + \mathbb{E}_\infty^\theta[\phi]\frac{1}{\mathbb{P}_\infty\{\psi \geq a\}}\\
&\quad + \mathbb{E}_\infty^\theta\left[T^c_{(b,0,\psi)}\big|\psi < b\right] \frac{1 - \mathbb{P}_\infty\{\psi \geq a\}}{\mathbb{P}_\infty\{\psi \geq a\}},\\
&\mathbb{E}_\infty^\theta[T_{(a,b,p)}]\\
&= \mathbb{E}_\infty^\theta[W_{(b,b,p,0)}] + \mathbb{E}_\infty^\theta[\phi]\frac{1}{\mathbb{P}_\infty\{\psi \geq a\}}\\
&\quad + \mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)}\big|\psi < b\right] \frac{1 - \mathbb{P}_\infty\{\psi \geq a\}}{\mathbb{P}_\infty\{\psi \geq a\}}.
\end{aligned}$$

Then

$$\begin{aligned}
&\mathbb{E}_\infty^\theta[T_{(a,b,p)}] - \mathbb{E}_\infty^\theta[T^c_{(a,b,0)}]\\
=&\mathbb{E}_\infty^\theta[W_{(b,b,p,0)}] - \mathbb{E}_\infty^\theta[T^c_{(b,0,0)}] + \Big(\mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)}\big|\psi < b\right]\\
&- \mathbb{E}_\infty^\theta\left[T^c_{(b,0,\psi)}\big|\psi < b\right]\Big) \frac{1 - \mathbb{P}_\infty\{\psi \geq a\}}{\mathbb{P}_\infty\{\psi \geq a\}}.
\end{aligned} \tag{21}$$

Given a and b, both $\mathbb{E}_\infty^\theta[W_{(b,b,p,0)}]$ and $\mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)}\big|\psi < b\right]$ monotonically go to infinity as $p \to 0$, while the other quantities in (21) are unrelated to p. Combining (18), we notice that to complete the proof, it suffices to prove the independence of $\hat{p}$ on a. Obviously, both $\mathbb{E}_\infty^\theta[W_{(b,b,p,0)}]$ and $\mathbb{E}_\infty^\theta[T^c_{(b,0,0)}]$ are unrelated with a. Also one can see that $\mathbb{E}_\infty^\theta\left[T^c_{(b,0,\psi)}\big|\psi < b\right] \leq \mathbb{E}_\infty^\theta[T^c_{(b,0,0)}]$. We then show in the following that $\mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)}\big|\psi < b\right]$ can be bounded below by a quantity that is independent of a.

$$\begin{aligned}
&\mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)}\big|\psi < b\right]\\
=&\mathbb{P}_\infty^\theta\{\psi = 0|\psi < b\}\mathbb{E}_\infty^\theta[W_{(b,b,p,0)}]\\
&+ \int_{0^+}^{b^-} \mathbb{E}_\infty^\theta\left[W_{(b,b,p,\omega)}\right] \mathrm{d}\mathbb{P}_\infty^\theta\{\psi \leq \omega\big|\psi < b\}\\
=&\mathbb{P}_\infty^\theta\{\psi = 0|\psi < b\}\mathbb{E}_\infty^\theta[W_{(b,b,1,0)}]\frac{1}{p} + \int_{0^+}^{b^-} \Big(\mathbb{E}_\infty^\theta\left[W_{(b,b,1,\omega)}\right]\\
&+ \mathbb{E}_\infty^\theta[W_{(b,b,1,0)}]\frac{1-p}{p}\Big) \mathrm{d}\mathbb{P}_\infty^\theta\{\psi \leq \omega\big|\psi < b\}\\
\geq&\mathbb{E}_\infty^\theta[W_{(b,b,1,0)}]\frac{1-p}{p},
\end{aligned} \tag{22}$$

where the second equality follows from Wald's identity. The proof thus is complete. ∎

As in Lemma 4, in the following lemma we prove that p can be chosen independent of a such that the sampling cost constraint is satisfied.

Lemma 5. *For any $c_X < \bar{c} \leq c_Y$ and $b > 0$, there exists $0 < \bar{p} < 1$ such that $\bar{p}$ is independent of a and $T_{(a,b,p)}$ with $p \in (0, \bar{p}]$ uniformly satisfies the average sampling cost* (2) *for any $a \geq b$.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

Proof: Let $\bar{\phi}$ be the limit value ϕ takes as a goes to infinity, i.e.,

$$\bar{\phi} = \inf\{i : \sum_{k=1}^{i} \ell(Z_k) < 0\}.$$

As in the proof of Lemma 4, by interpreting the algorithm as a sequence of two-sided SPRTs, one can see that for any $T_{(a,b,p)}$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_\infty^\theta\left[\sum_{k=1}^{n} c_Z \middle| T_{(a,b,p)} > n\right] \\
\leq &\limsup_{n\to\infty} \frac{1}{n - T_{(b,b,p)}}\mathbb{E}_\infty^\theta\left[\sum_{k=T_{(b,b,p)}+1}^{n} c_Z \middle| T_{(a,b,p)} > n\right] \\
= &\frac{c_X \mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)} \middle| \psi < b\right] + c_Y \mathbb{E}_\infty^\theta\left[\phi \middle| \psi < b\right]}{\mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)} \middle| \psi < b\right] + \mathbb{E}_\infty^\theta\left[\phi \middle| \psi < b\right]} \\
\leq &\frac{c_X \mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)} \middle| \psi < b\right] + c_Y \mathbb{E}_\infty^\theta\left[\bar{\phi}\right]}{\mathbb{E}_\infty^\theta\left[W_{(b,b,p,\psi)} \middle| \psi < b\right] + \mathbb{E}_\infty^\theta\left[\bar{\phi}\right]} \\
:= &\check{c}(\theta)
\end{aligned}$$

where the first inequality follows from the fact that $\gamma_k = 0, \forall k \in [1, T_{(b,b,p)}]$; the first equality follows from the alternating renewal process theory (Page 173, [30]); the second inequality follows from the fact that $\check{c}(\theta)$ is monotonically increasing with $\mathbb{E}_\infty^\theta\left[\phi \middle| \psi < b\right]$ and $\mathbb{E}_\infty^\theta\left[\phi \middle| \psi < b\right] \uparrow \mathbb{E}_\infty^\theta[\bar{\phi}]$ as $a \to \infty$. It is well known that $\mathbb{E}[\bar{\phi}] < \infty$ (e.g., Corollary 2.4, [31]). Then by (22), one completes the proof. ■

We shall prove Theorem 2 now. It is well known that $\mathbb{E}_\infty^\theta[T_{(\ln\zeta,0,p)}] > \zeta$ (e.g., Lemma 1, [27]). By Lemma 4 and 5, to satisfy the constraint (4) and (5), one can let $p^* = \min(\bar{p}, \hat{p})$. We then focus on (7). To highlight $a = \ln\zeta$, with a little abuse of notation, in the remainder of this proof we rewrite ϕ in (19) and ψ in (20) as $\phi(\ln\zeta)$ and $\psi(\ln\zeta)$, respectively. Since

$$\mathbb{E}_1^\theta\left[W_{(b,b,p,\psi(\ln\zeta))} \middle| \psi(\ln\zeta) < b\right] \leq \mathbb{E}_1^\theta[T_{(b,b,p)}]$$

and given b and p, as $\zeta \to \infty$

$$\frac{\mathbb{E}_1^\theta[T_{(b,b,p)}]}{\mathbb{E}_1^\theta[\phi(\ln\zeta)]} \to 0,$$

one obtains that

$$\begin{aligned}
&\mathbb{E}_1^\theta[T_{(\ln\zeta,b,p)}] \\
&= \mathbb{E}_1^\theta[T_{(b,b,p)}] + \mathbb{E}_1^\theta[\phi(\ln\zeta)]\frac{1}{\mathbb{P}_1\{\psi(\ln\zeta) \geq \ln\zeta\}} \\
&\quad + \mathbb{E}_1^\theta\left[W_{(b,b,p,\psi(\ln\zeta))} \middle| \psi(\ln\zeta) < b\right]\frac{1 - \mathbb{P}_1\{\psi(\ln\zeta) \geq \ln\zeta\}}{\mathbb{P}_1\{\psi(\ln\zeta) \geq \ln\zeta\}} \\
&= \mathbb{E}_1^\theta[\phi(\ln\zeta)]\frac{1}{\mathbb{P}_1\{\psi(\ln\zeta) \geq \ln\zeta\}}(1 + o(1)), \quad \text{as } \zeta \to \infty \\
&\leq \frac{\ln\zeta - b}{\mathbf{I}(g_Y||f_Y)}(1 + o(1)), \quad \text{as } \zeta \to \infty \\
&\leq \frac{\ln\zeta}{\mathbf{I}(g_Y||f_Y)}(1 + o(1)), \quad \text{as } \zeta \to \infty,
\end{aligned}$$

where the first inequality follows from Wald's identity and the well established properties of the CuSum algorithm (Page 142 and 159, [10]). The proof is complete.

Appendix C
Proof of Theorem 3

A. *Proof of equation* (13)

In this section, we focus on the proof of equation (13). To this end, we define the following:

$$\begin{aligned}
K^* &= \sup\{k : \exists m \in \Xi \text{ such that } \gamma_{m,k} = 0\}, \\
K_m^* &= \sup\{k : \gamma_{m,k} = 0\}, m \in \Xi.
\end{aligned}$$

Note that after the time instant K^*, *all* the *affected* sensor nodes consistently use the expensive observations $\{Y_{m,k}\}$. The term K_m^* has a similar meaning, but only for the sensor m. In the following lemma, we show that for the MOS-CuSum algorithm $T_{(a,b,p)}$, when a goes to infinity, the expected value of K^* remains finite.

Lemma 6. *For any non-empty Ξ, $\nu < \infty$ and the MOS-CuSum algorithm $T_{(a,b,p)}$ with any finite b and $0 \prec p \preceq 1$, as $a \to \infty$, there holds*

$$\mathbb{E}_\nu^{\Theta,\Xi}[K^*] < \infty.$$

Proof: Note that since

$$\begin{aligned}
\mathbb{E}_\nu^{\Theta,\Xi}[K^*] &= \nu + \mathbb{E}_\nu^{\Theta,\Xi}[\max_{m\in\Xi}(K_m^* - \nu)] \\
&\leq \nu + \sum_{m\in\Xi}\mathbb{E}_\nu^{\Theta,\Xi}[K_m^* - \nu],
\end{aligned}$$

it suffices to prove that as $a \to \infty$,

$$\mathbb{E}_\nu^{\Theta,\Xi}[K_m^* - \nu] < \infty, \quad \forall m \in \Xi. \tag{23}$$

Just like the $k_Y(i)$ defined in the proof of Lemma 2, for the sensor $m \in \Xi$, we "pick" the time instants the expensive observations $\{Y_{m,k}\}$ are sampled after time ν by the following:

$$\begin{aligned}
&\mathcal{I}_m(1) = \inf\{k : k > \nu, \gamma_{m,k} = 1\}, \\
&\mathcal{I}_m(i) = \inf\{k : k > \mathcal{I}_m(i-1), \gamma_{m,k} = 1\}, \ \forall i \geq 2.
\end{aligned}$$

We then define $\varphi_{m,i}$ as the statistic of the CuSum algorithm of these expensive observations, i.e.,

$$\varphi_{m,i} = (\varphi_{m,i-1} + \ell(Z_{m,\mathcal{I}_m(i)}))^+, \ i \geq 1$$

with $\varphi_{m,0} = 0$. Then we define

$$I_m^* = \sup\{i : \varphi_{m,i} \leq 0\}, m \in \Xi.$$

Then by Theorem D in [32] and the finiteness assumption of the second moment of the K–L divergences (8), using the path-wise arguments, one obtains that

$$\mathbb{E}_\nu^{\Theta,\Xi}[I_m^*] < \infty.$$

Note that for K_m^*, the worst case is that after ν, the detection statistic $s_{m,k}$ goes below b_m each time when it crosses b_m from below and updates with *only one* expensive observation $Y_{m,k}$. Then by Wald's identity one obtains that

$$\begin{aligned}
\mathbb{E}_\nu^{\Theta,\Xi}[K_m^* - \nu] &\leq \mathbb{E}_1^\theta\left[W_{(b_m,b_m,p_m,\psi)} \middle| \psi < b_m\right]\mathbb{E}_\nu^{\Theta,\Xi}[I_m^*] \\
&\leq \mathbb{E}_1^\theta\left[W_{(b_m,b_m,p_m,0)}\right]\mathbb{E}_\nu^{\Theta,\Xi}[I_m^*] \\
&< \infty.
\end{aligned}$$

Recall that $W_{(a,b,p,\omega)}$ is defined in the proof of Lemma 4. The proof thus is complete. ■

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2016.2609998, IEEE Transactions on Automatic Control

To prove (13), we define the following stopping time

$$\hat{T} = \inf\{k : \hat{s}_k \geq a\},$$

where

$$\hat{s}_k = \sum_{i=1}^{k} \sum_{m \in \Xi} \ln \frac{g_{Y,m}(Z_{m,i+K^*})}{f_{Y,m}(Z_{m,i+K^*})} + \sum_{m \in \Xi} b_m, \forall k \geq 1.$$

From the definition of $s_{m,k}$ and K^*, one can see that $\forall k \geq 1$, $\hat{s}_k$ is dominated by $\sum_{m \in \Xi} s_{m,k+K^*}$. What is more, since $\sum_{m \in \Xi} s_{m,T(a,b,p)} \leq a$, the following holds:

$$\mathbb{E}_{\nu}^{\Theta,\Xi}[T(a,b,p) - K^*] \leq \mathbb{E}_{1}^{\Theta,\Xi}[\hat{T}]. \tag{24}$$

By renewal theory (Page 168, [28]), as $a \to \infty$,

$$\mathbb{E}_{1}^{\Theta,\Xi}[\hat{T}] = \frac{a - \sum_{m \in \Xi} b_m}{\sum_{m \in \Xi} \mathbf{I}(g_{Y,m}||f_{Y,m})} + O(1). \tag{25}$$

Thus, by Lemma 6, as $a \to \infty$,

$$\begin{aligned}\mathbb{E}_{\nu}^{\Theta,\Xi}[T(a,b,p)] =& \mathbb{E}_{\nu}^{\Theta,\Xi}[T(a,b,p) - K^*] + \mathbb{E}_{\nu}^{\Theta,\Xi}[K^*] \\ \leq& \frac{a}{\sum_{m \in \Xi} \mathbf{I}(g_{Y,m}||f_{Y,m})}(1 + o(1)).\end{aligned}$$

The proof thus is complete.

B. *Proof of equation* (14)

In this section, we focus on the proof of (14), which is done using the same reasoning as in the proof of Theorem 1 of [20]. The difference is that the observations at each sensor node in our case ceases to be i.i.d. conditioned on the change event. We thus provide the following two lemmas.

Lemma 7. *For any* $m \in \mathcal{M}$, $k \in \mathbb{N}_+$ *and any* $a \in \mathbb{R}_+$,

$$\mathbb{P}_{\infty}^{\Theta,\Xi}\{s_{m,k} \geq a\} \leq e^{-a}.$$

Proof: Define the quantity $\hat{s}_{m,k}$ as follows:

$$\begin{aligned}&\hat{s}_{m,0} = 0, \\ &\hat{s}_{m,k} = (\hat{s}_{m,k-1} + \ell(Z_{m,k}))^+,\end{aligned}$$

where the observation scheduling policy is the same as defined in (12). The difference between $\hat{s}_{m,k}$ and $s_{m,k}$ is that $\hat{s}_{m,k}$ crosses b_m deterministically and $\hat{s}_{m,k}$ has no reset action when crossing b_m from below. Since $\hat{s}_{m,k} \geq s_{m,k}, \forall k \in \mathbb{N}_+$, then for any $a \geq 0$

$$\mathbb{P}_{\infty}^{\Theta,\Xi}\{s_{m,k} \geq a\} \leq \mathbb{P}_{\infty}^{\Theta,\Xi}\{\hat{s}_{m,k} \geq a\}.$$

Given any scheduling realization for $s_{m,k}, \gamma_1, \dots, \gamma_k$, define the following quantities:

$$\hat{Z}_{m,i} = (1 - \gamma_{k+1-i})X_{m,i} + \gamma_{k+1-i}Y_{m,i}, \forall 1 \leq i \leq k$$

and $\breve{s}_{m,k} = \sum_{i=1}^{k} \hat{Z}_{m,i}$. Then it is well known that (see Appendix 2 of [28] or Lemma 3 of [27]) that $\hat{s}_{m,k}$ and $\max_{1 \leq i \leq k} \breve{s}_{m,i}$ have the same distribution and

$$\begin{aligned}\mathbb{P}_{\infty}^{\Theta,\Xi}\{\hat{s}_{m,k} \geq a\} =& \mathbb{P}_{\infty}^{\Theta,\Xi}\{\max_{1 \leq i \leq k} \breve{s}_{m,k} \geq a\} \\ =& \mathbb{P}_{\infty}^{\Theta,\Xi}\{\tau_m(a) \leq k\} \\ \leq& \mathbb{P}_{\infty}^{\Theta,\Xi}\{\tau_m(a) \leq \infty\} \\ \leq& e^{-a},\end{aligned}$$

where $\tau_m(a) = \inf\{k : \breve{s}_{m,k} \geq a\}$ and the last inequality follows from the fact that Wald's likelihood identity (Page 10, [28]) applies even if the observations are not i.i.d. The proof thus is complete. ■

Lemma 8. *Let* $\tau = \inf\{k : s_{m,k} = 0, \forall m \in \mathcal{M}\}$ *be the first time that the local statistic of each sensor reaches zero simultaneously. Then* $\mathbb{E}_{\infty}^{\Theta,\Xi}[\tau] < \infty$.

Proof: Let

$$\mathfrak{p}_m^* = \mathbb{P}_{\infty}^{\Theta,\Xi}\{s_{m,k} = 0\}, \quad \text{as } k \to \infty.$$

Then as in the proof of Proposition 1 of [20], by renewal theory one can conclude that

$$\mathbb{E}_{\infty}^{\Theta,\Xi}[\tau] = \frac{1}{\prod_{m=1}^{M} \mathfrak{p}_m^*}.$$

It suffices to prove that $\mathfrak{p}_m^* > 0, \forall m \in \{1, \dots, M\}$. Note that since the observations are not i.i.d., the relation (A2) in the proof of Proposition 1 of [20] does not hold. We prove $\mathfrak{p}_m^* > 0$ by treating $s_{m,k}$ as a homogenous Markov process. Let $\mathbb{P}^*$ be the equilibrium distribution of $s_{m,k}$. Since $\mathbf{I}(f_{X,m}||g_{X,m}) > 0$, there must exist $0 < x^* < b_m$ such that $\mathbb{P}_{\infty}^{\Theta,\Xi}\{\ell(X_k) \leq -x^*\} > 0$. Then

$$\begin{aligned}\mathfrak{p}_m^* >& \int_0^{x^*} \mathbb{P}_{\infty}^{\Theta,\Xi}\{\ell(x_k) \leq -s_{m,k-1}\} \mathrm{d}\mathbb{P}^*(s_{m,k-1}) \\ \geq& \mathbb{P}_{\infty}^{\Theta,\Xi}\{\ell(x_k) \leq -x^*\} \int_0^{x^*} \mathrm{d}\mathbb{P}^*(s_{m,k-1}) \\ \geq& \mathbb{P}_{\infty}^{\Theta,\Xi}\{\ell(x_k) \leq -x^*\}(1 - e^{-x^*}) \\ >& 0,\end{aligned}$$

where the second last inequality follows from Lemma 7. The proof thus is complete. ■

The outline of the proof is as follows. By Lemma 7, as in Lemma B1 of [20] one can easily obtain that for any $k \in \mathbb{N}_+$ and $a \geq 0$, $\mathbb{P}_{\infty}^{\Theta,\Xi}\{\sum_{m=1}^{M} s_{m,k} \geq a\} \leq e^{-a} \sum_{m=0}^{M-1} \frac{a^m}{m!}$. As in Lemma B2 of [20], the whole time horizon of our algorithm can be broken into subintervals (renewed when $s_{m,k} = 0, \forall m \in \{1, \dots, M\}$), between which $s_{m,k}$ are i.i.d. Combining Lemma 8, one can obtain that for any $t > 0$, $\liminf_{a \to \infty} \mathbb{P}_{\infty}^{\Theta,\Xi}\{T(a,b,p) \geq t \frac{e^a}{\sum_{m=0}^{M-1} a^m/m!}\} \geq e^{-t}$. Then equation (14) follows using the same reasoning as in Theorem 1 of [20].

References

[1] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in minimax settings," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6917–6931, Oct 2013.

[2] ——, "Data-efficient quickest change detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3727–3735, 2015.

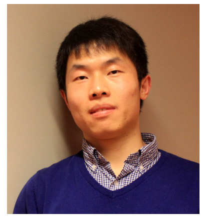
[3] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Transactions on Automatic Control,*, vol. 58, no. 10, pp. 2451–2464, 2013.

[4] J. Whitehead, *The design and analysis of sequential clinical trials*. John Wiley & Sons, 1997.

[5] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013.

[6] Q. Maillet, H. Xu, N. Ozay, and R. M. Murray, "Dynamic state estimation in distributed aircraft electric control systems via adaptive submodularity," in *Proceedings of the 52nd Annual Conference on Decision and Control (CDC)*, 2013, pp. 5497–5503.

[7] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. ACM, 2002, pp. 88–97.
[8] B. O'Flyrm, R. Martinez, J. Cleary, C. Slater, F. Regan, D. Diamond, and H. Murphy, "Smartcoast: a wireless sensor network for water quality monitoring," in *32nd IEEE Conference on Local Computer Networks*, 2007, pp. 815–816.
[9] A. N. Shiryaev, *Optimal stopping rules*. New York: Springer, 2007.
[10] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge: Cambridge University Press, 2009.
[11] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection with on–off observation control," *Sequential Analysis*, vol. 31, no. 1, pp. 40–77, 2012.
[12] J. Geng and L. Lai, "Non-bayesian quickest change detection with stochastic sample right constraints," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5090–5102, Oct 2013.
[13] T. Banerjee and V. V. Veeravalli, "Data-efficient minimax quickest change detection in a decentralized system," *Sequential Analysis*, vol. 34, no. 2, pp. 148–170, 2015.
[14] P. Karumbu and A. Kumar, "Optimum sleep-wake scheduling of sensors for quickest event detection in small extent wireless sensor networks," *arXiv preprint arXiv:1105.6024*, 2011.
[15] A. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability and Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.
[16] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.
[17] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, pp. 206–227, 1985.
[18] V. Srivastava, F. Pasqualetti, and F. Bullo, "Stochastic surveillance strategies for spatial quickest detection," *The International Journal of Robotics Research*, vol. 32, no. 12, pp. 1438–1458, 2013.
[19] E. Page, "Continuous inspection schemes," *Biometrika*, pp. 100–115, 1954.
[20] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.
[21] ——, "Quickest detection in censoring sensor networks," in *Proceedings of International Symposium on Information Theory Proceedings (ISIT)*, 2011, pp. 2148–2152.
[22] X. Ren, K. H. Johansson, D. Shi, and L. Shi, "Quickest change detection in adaptive censoring sensor networks," *arXiv preprint arXiv:1503.04999*, 2015.
[23] V. Srivastava, K. Plarre, and F. Bullo, "Randomized sensor selection in sequential hypothesis testing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2342–2354, 2011.
[24] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4938–4953, 2011.
[25] V. Krishnamurthy, "How to schedule measurements of a noisy Markov chain in decision making?" *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4440–4461, 2013.
[26] T. M. Cover and J. A. Thomas, *Elements of information theory*. New Jersey: John Wiley & Sons, 2006.
[27] Y. Mei, "Information bounds and quickest change detection in decentralized decision systems," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2669–2681, 2005.
[28] D. Siegmund, *Sequential analysis: tests and confidence intervals*. New York: Springer, 1985.
[29] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.
[30] S. Asmussen, *Applied probability and queues*. New York: Springer, 2003, vol. 51.
[31] M. Woodroofe, *Nonlinear renewal theory in sequential analysis*. Philadelphia, PA: Siam, 1982, vol. 39.
[32] J. Kiefer and J. Sacks, "Asymptotically optimum sequential inference and design," *The Annals of Mathematical Statistics*, pp. 705–750, 1963.



Xiaoqiang Ren received the B.E. degree in the Department of Control Science and Engineering from Zhejiang University, Hangzhou, China, in 2012 and the Ph.D. degree in the Department of Electronic and Computer Engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2016. He is currently a research associate in the Department of Electronic and Computer Engineering at Hong Kong University of Science and Technology. From January 2015 to June 2015, he was a visiting student in the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Sweden. From August 2015 to January 2016, he was a visiting student in NeSC group, Zhejiang University. His research interests include sequential detection, security of cyber-physical system and networked control and estimation.

Karl H. Johansson is Director of the Stockholm Strategic Research Area ICT The Next Generation and Professor at the School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. He received MSc and PhD degrees in Electrical Engineering from Lund University. He has held visiting positions at UC Berkeley, California Institute of Technology, Nanyang Technological University, Institute of Advanced Studies Hong Kong University of Science and Technology, and Norwegian University of Science and Technology. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation systems. He is a member of the IEEE Control Systems Society Board of Governors and the European Control Association Council. He is past Chair of the IFAC Technical Committee on Networked Systems. He has been on the Editorial Boards of Automatica, IEEE Transactions on Automatic Control, and IET Control Theory and Applications. He is currently a Senior Editor of IEEE Transactions on Control of Network Systems and Associate Editor of European Journal of Control. He was the General Chair of the ACM/IEEE Cyber-Physical Systems Week 2010 in Stockholm and IPC Chair of many conferences. He received the Best Application Paper Award of IEEE Transactions on Automation Science and Engineering 2015, the Best Paper Award of the IEEE International Conference on Mobile Ad-hoc and Sensor Systems 2009 and the Best Theory Paper Award of the World Congress on Intelligent Control and Automation 2014. In 2009 he was awarded Wallenberg Scholar, as one of the first ten scholars from all sciences, by the Knut and Alice Wallenberg Foundation. He has held a Senior Researcher Position with the Swedish Research Council. He was awarded Future Research Leader from the Swedish Foundation for Strategic Research in 2005. He received the triennial Young Author Prize from IFAC in 1996 and the Peccei Award from the International Institute of System Analysis, Austria, in 1993. He received Young Researcher Awards from Scania in 1996 and from Ericsson in 1998 and 1999. He is a Fellow of the IEEE.

Ling Shi received the B.S. degree in electrical and electronic engineering from Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2002 and the Ph.D. degree in Control and Dynamical Systems from California Institute of Technology, Pasadena, CA, USA, in 2008. He is currently an associate professor at the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include cyber-physical systems security, networked control systems, sensor scheduling, and event-based state estimation. He has been serving as a subject editor for International Journal of Robust and Nonlinear Control from March 2015 and an associate editor for IEEE Transactions on Control of Network Systems from July 2016. He also served as an associate editor for a special issue on Secure Control of Cyber Physical Systems in the IEEE Transactions on Control of Network Systems in 2015-2016.