Quickest Change Detection in Adaptive Censoring Sensor Networks

Xiaoqiang Ren, Karl H. Johansson, *Fellow, IEEE*, Dawei Shi, and Ling Shi, *Member, IEEE*

Abstract—The problem of quickest change detection with communication rate constraints is studied. A network of wireless sensors with limited computation capability monitors the environment and sends observations to a fusion center via wireless channels. At an unknown time instant, the distributions of observations at all the sensor nodes change simultaneously. Due to limited energy, the sensors cannot transmit at all the time instants. The objective is to detect the change at the fusion center as quickly as possible, subject to constraints on false detection and average communication rate between the sensors and the fusion center. A minimax formulation is proposed. The cumulative sum (CuSum) algorithm is used at the fusion center and censoring strategies are used at the sensor nodes. The censoring strategies, which are adaptive to the CuSum statistic, are fed back by the fusion center. The sensors only send observations that fall into prescribed sets to the fusion center. This CuSum adaptive censoring (CuSum-AC) algorithm is proved to be an equalizer rule and to be globally asymptotically optimal for any positive communication rate constraint, as the average run length to false alarm goes to infinity. It is also shown, by numerical examples, that the CuSum-AC algorithm provides a suitable trade-off between the detection performance and the communication rate.

Index Terms—Adaptive, asymptotically optimal, censoring, CuSum, minimax, quickest change detection, wireless sensor networks.

I. Introduction

Background and Motivations: the goal of quickest change detection is to detect the abrupt change in stochastic processes as quickly as possible subject to certain constraints on false detection. This problem has a wide range of applications, such as habitat monitoring [1], quality control engineering [2], computer security [3] and cognitive radio networks [4]. In the classical quickest change detection formulation, the decision maker observes a sequence of observations $\{X_1, \ldots, X_k, \ldots\}$, the distribution of which changes at an unknown time instant ν. The observations before the change $X_1, \ldots, X_{\nu-1}$ are independent and identically distributed (i.i.d.), and the observations after the change $X_\nu, \ldots, X_\infty$ are also i.i.d. but with a different distribution. The change event model distinguishes two problem formulations: the Bayesian formulation due to Shiryaev [5], [6] and the minimax formulation due to Lorden [7] and Pollak [8].

The classical quickest change detection problem does not consider the cost of acquiring observations. It assumes that the decision maker can access observations at all the time instants freely. This is an issue for resource-limited applications, such as those using wireless sensor networks (WSNs). In the problem of quickest change detection with WSNs, observations are taken by one or multiple sensors, which communicate with the decision maker via wireless channels [9]–[11]. The limited resources, which include limited energy for each battery-powered sensor and the limited communication bandwidth, naturally pose the constraint that the observations cannot be sent to the decision maker continuously. Thus, we consider the problem of quickest change detection with such constraints.

Related Literatures and Contributions: Recently, there are several works on constrained quickest change detection with minimax formulations [11]–[14]. Two classes of characterizations of the cost acquiring observations were considered: the cost of sampling [11], [13], [14] and the cost of communication [10], [12], [14]. In these works, algorithms consisting of stopping times and sampling/transmission schedulers were proposed. By transmission (sampling) schedulers, when local sensors send their data to a fusion center (a decision maker samples) is determined. The communication constraint was studied from the perspective of quantization as well as the communication rate in [15] but with general sequential detection settings, where an interesting point was that only one bit of information was sent to the fusion center whenever a transmission occured.

In this paper, we only take the cost of communication into account. This is motivated by WSNs applications for which the energy cost of sampling is usually negligible compared with that of communication [16], [17]. Thus, it is a reasonable assumption that the sensors can take observations at each time instant but with limited number of communications with the fusion center. Furthermore, as in [10], [12], and [14], we do not consider quantization errors of the data sent from local sensors to the fusion center.

The structure of the system considered in this paper is illustrated in Fig. 1. Observations are taken by M sensors and are sent to a fusion center via wireless channels. Due to limited energy, the remote sensors cannot transmit at all the
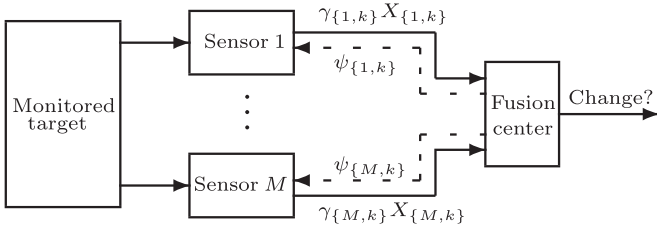
Fig. 1. Quickest change detection system in adaptive censoring sensor networks. Each sensor corresponds to the blue rectangle in Fig. 2.

Fig. 2. Quickest change detection system with a sensor that adopts an adaptive censoring strategy.

time instants. To make the best use of the limited resources, the sensors are assumed to adopt censoring strategies [18]. Each of the sensors samples at each time instant, but only transmits informative samples. The censoring strategies are adaptive to the detection statistic available at the fusion center. When necessary, the fusion center tells the sensors about the censoring strategies to use via the feedback channels.[1]

To deal with communication constraints, in the existing literature [10], [12], [14], [19], CuSum-like algorithms (CuSum in [10] and [12], a variant called DE-CuSum in [14] and [19]) are run locally at the sensor nodes and *the detection statistic of the algorithm* is sent to the fusion center only when it is above a certain threshold. In this paper, a fundamentally different approach is adopted: *the observations* instead of the detection statistic are censored and transmitted to the fusion center. Compared to running CuSum-like algorithms, the online computation load required at the sensor side to censor the observations is reduced (see Remark 1). Another advantage is that our algorithm is an equalizer rule, which helps reduce off-line computation complexity (see Remark 2). Compared with decentralized settings in [10], [12], [14], and [19], where the communication is unidirectional and remote sensors implement the censoring strategy in an autonomous manner, our algorithm is centralized in the sense that the censoring strategies used at remote sensors are fed back by the fusion center. Evidently, the feedback transmission introduced complicates the system, although it occurs only occasionally and its message is rather simple. Our algorithm, however, is able to reduce system complexity by reducing the number of sensors required. More specifically, to achieve the same detection performance, fewer sensors are required compared with the decentralized counterparts. A similar idea can be found in [20]. About the decentralized censoring strategies, we point out that to utilize the information of past observations available at each sensor, it is necessary to censor the detection statistic instead of current observations. Note that in this paper, we do not consider the cost of sampling, while if the sampling is "energy consuming" (i.e., the energy consumed by sampling is comparable to that by communication), then the DE-CuSum algorithm running locally at the sensor nodes [14], which skips sampling when necessary, is desirable.

In summary, the main contribution is that for the quickest change detection with communication rate constraints in the

minimax formulation, a novel algorithm is proposed, which is the CuSum algorithm coupled with adaptive censoring strategies (CuSum-AC). The CuSum-AC algorithm is proven to be asymptotically optimal for any positive communication rate constraint and thus provides some insights into how one can utilize censoring strategies *adaptively* to achieve the globally asymptotic optimality.

Organizations: The remainder of this paper is organized as follows. The one sensor case is studied in Sections II–IV. The mathematical formulation of the considered problem is given in Section II. We present the CuSum-AC algorithm in Section III. The main results are given in Section IV. First we show that the CuSum-AC algorithm is an equalizer rule, i.e., the worst-worst case detection delay is attained whenever the change event happens. Then we prove that the CuSum-AC algorithm is globally asymptotically optimal for any positive communication rate constraint. We generalize the results obtained for the one sensor case to the multiple sensors scenario in Section V. Numerical examples are given in Section VI to illustrate the main results. Some concluding remarks are given in the end. All the proofs are presented in Appendices.

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. $k \in \mathbb{N}$ is the time index. $\mathbf{1}_A$ represents the indicator function that takes value 1 on the set A and 0 otherwise. $\times$ stands for the Cartesian product. For $x \in \mathbb{R}$, $(x)^+ = \max(0, x)$.

II. PROBLEM SETUP

For simplicity of presentation, first we consider the one sensor case; see Fig. 2. Then we extend the results to the sensor networks scenario in Section V. Consider the change detection system in Fig. 2. A sequence of observations, say $\{X_k\}_{k \in \mathbb{N}_+}$, about the monitored environment are taken locally at the sensor. Assume that ν is an unknown (but not random) time instant when a change event takes place. The instant may be ∞, corresponding to that the change never happens. The observations at the sensor before ν, $\{X_1, \ldots, X_{\nu-1}\}$, are i.i.d. with probability density function (pdf) f_0, and the observations from ν on, $X_\nu, X_{\nu+1}, \ldots$, are i.i.d. with pdf f_1. Let $\mathbb{P}_\nu$ denote the probability measure when the change happens at ν. If there is no change, we denote this measure by $\mathbb{P}_\infty$. The expectation $\mathbb{E}_\nu$ and $\mathbb{E}_\infty$ are defined accordingly.

To characterize the behavior that the sensor cannot send the observation X_k to the decision maker all the time, we introduce a binary variable γ_k as

$$\gamma_k = \begin{cases} 1, & \text{if } X_k \text{ is sent to the decision maker} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

---

[1]The feedback transmissions may cause additional energy consumption at the sensor side. But one should note that in our algorithm, the feedback message is quite simple (see Remark 1) and the feedback transmissions are usually quite few (see Remark 5). In particular, when CuSum-AC algorithm with $N = 2$ levels is used, only one bit of information is needed when feedback occurs.
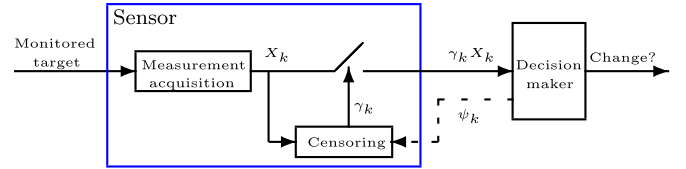
Then the information pattern available for the decision maker at the time instant k is given by $\mathcal{I}_k = \{(1, \gamma_1, \gamma_1 X_1), \ldots, (k, \gamma_k, \gamma_k X_k)\}$ with $\mathcal{I}_0 = \{\emptyset\}$. A random variable $T \in \mathbb{N}_+$ is called a stopping time if $\{T = k\} \in \sigma(\mathcal{I}_k)$, where $\sigma(\mathcal{I}_k)$ is the smallest sigma-algebra of $\mathcal{I}_k$. A stopping time can be characterised by a stopping rule, which is a mechanism that decides whether or not to stop based on the available information.

To make the best use of the limited communication resources, the censoring strategy is implemented at the sensor node. We consider an adaptive censoring strategy, which varies with the information pattern. Specifically, the censoring strategy used at the sensor node at time instant k, which is denoted by ψ_k, is determined by the decision maker based on $\mathcal{I}_{k-1}$. When $\psi_k \neq \psi_{k-1}$, the decision maker sends ψ_k to the sensor through the feedback channel. Since the sensor is assumed to have no memory and can thus only access X_k at time k, the censoring strategy $\psi_k : \mathbb{R} \mapsto \{0, 1\}$ has the form as $\gamma_k = \psi_k(X_k)$. The censoring policy along the horizon is given by $\Psi = (\psi_1, \ldots, \psi_T)$.

The communication constraint is formulated as the limited communication rate *before* the change event happens. It depends on the censoring policy Ψ and is formalized as

$$r(\Psi) = \limsup_{n \to \infty} \frac{1}{n} \mathbb{E}_\infty \left[\sum_{k=1}^n \gamma_k \,\middle|\, T \geq n \right] \leq \epsilon \qquad (2)$$

where $0 < \epsilon \leq 1$ is a design parameter. By adjusting ϵ, a tradeoff between communication resources and detection performance is obtained. Note that the post-change period (i.e., the detection delay) is usually quite small compared with the pre-change period, hence we only pose the communication constraint *before* the change. The conditional expectation $\mathbb{E}_\infty[\cdot | T \geq n]$ thus is considered. The asymptotic optimality result of the paper does not hold if the total cost is considered (see Remark 3). A similar criterion called pre-change transmission cost is considered in [19].

For the detection performance of the quickest change detection, there are two indices: the risk of false detection and the detection delay. Given T and Ψ, the risk of false detection is characterized by the average run length to false alarm (ARLFA)

$$g(T, \Psi) = \mathbb{E}_\infty[T]$$

cf., [7], [21]. Note that the reciprocal of the ARLFA is connected to the false alarm rate. When the ARLFA goes to infinity, the false alarm rate goes to zero. The stopping time T is related to $\mathcal{I}_k$, which is determined by the observation sequence $\{X_k\}$ and the censoring policy Ψ, so the ARLFA is related with Ψ. To highlight this dependence, we use $g(T, \Psi)$ in the above definition.

For the detection delay, we consider Lorden's worst-worst case detection delay [7],[2] which is given by

$$d_L(T, \Psi) = \sup_{1 \leq \nu < \infty} \left\{ \operatorname*{ess\,sup}_{\mathcal{I}_{\nu-1}} \mathbb{E}_\nu \left[(T - \nu + 1)^+ | \mathcal{I}_{\nu-1} \right] \right\}. \quad (3)$$

---

[2]It is easy to see that the main result in this paper, i.e., asymptotical optimality of our algorithm, also holds when Pollak's criterion [8] is considered.

Problem 1:

$$\begin{aligned} \underset{T, \Psi}{\text{minimize}} \quad & d_L(T, \Psi) \\ \text{subject to} \quad & g(T, \Psi) \geq \zeta & (4) \\ & r(\Psi) \leq \epsilon & (5) \end{aligned}$$

where $\zeta \geq 1$ is a given lower bound of the ARLFA.

Note that for the classical formulation of the quickest change detection, the observations are assumed to be i.i.d. conditioned on the change event. While since Ψ is adaptive, the available observation sequence $\{\gamma_k, \gamma_k X_k\}$ are correlated across the time. To solve the above problems thus is quite challenging.

To avoid degenerate problems, we make the following assumption for the remainder of this paper.

Assumption 1:

$$0 < \mathbf{I}(f_1 \| f_0) < \infty, \quad 0 < \mathbf{I}(f_0 \| f_1) < \infty$$

where $\mathbf{I}(f_1 \| f_0) = \int_{\mathbb{R}} f_1(x) \ln(f_1(x)/f_0(x)) \mathrm{d}x$, $\mathbf{I}(f_0 \| f_1) = \int_{\mathbb{R}} f_0(x) \ln(f_0(x)/f_1(x)) \mathrm{d}x$ are the Kullback-Leibler (K-L) divergences.

Our subsequent analysis utilizes the CuSum algorithm, which is stated as follows. Let constant a be a given threshold and $\ell(X_k) = \ln(f_1(X_k)/f_0(X_k))$ the log-likelihood ratio function. The stopping time for the CuSum algorithm thus is computed as

$$T(a) = \inf\{k : c_k > a\} \qquad (6)$$

where c_k is the detection statistic for the CuSum algorithm computed by

$$c_k = (c_{k-1} + \ell(X_k))^+$$

with $c_0 = 0$. The CuSum algorithm is optimal for original Lorden's formulation when a is chosen such that $\mathbb{E}_\infty[T(a)] = \zeta$ [22], [23]. When there is no communication rate constraint, i.e., $\epsilon = 1$, Problem 1 is reduced to original Lorden's formulation. We should remark that it is difficult to find strictly optimal algorithms for Problem 1. We hence focus on asymptotically optimal solution. For simplicity, we use $T(a)$ to denote the CuSum algorithm with the threshold a for the remainder of this paper.

III. CuSum-AC Algorithm

In this section, we present the proposed CuSum-AC algorithm, which is the CuSum algorithm coupled with a censoring policy that adaptively switches between different censoring strategies. We say a CuSum-AC algorithm is with N levels if the number of censoring strategies used is N. For ease of presentation, we present the CuSum-AC algorithm with $N = 2$ levels, and for the CuSum algorithm with $N > 2$ levels, see (12) and Remark 4. In the remainder of this paper, if not particularly indicated, the CuSum-AC algorithm refers to the one with $N = 2$ levels. The algorithm consists of three parts: how the detection statistic updates, what the adaptive censoring policy is, and when the algorithm stops and declares the change.

Let a and a_1 be given thresholds with $a_1 < a$. The detection statistic s_k are updated as follows:

$$\tilde{s}_k = \max\left\{0, s_{k-1} + \ell^{\psi_k}(\gamma_k, X_k)\right\} \qquad (7)$$

$$s_k = \begin{cases} a_1, & \text{if} \quad s_{k-1} < a_1 \text{ and } \tilde{s}_k \geq a_1 \\ \tilde{s}_k, & \text{otherwise} \end{cases} \qquad (8)$$

with initial value $s_0 = 0$. The quantity $\ell^{\psi_k}(\gamma_k, X_k)$ is the log-likelihood ratio function of the random variable $\gamma_k X_k$[3] under the censoring strategy ψ_k

$$\ell^{\psi_k}(\gamma_k, X_k) = \ln \frac{f_1(X_k)}{f_0(X_k)}(\gamma_k - 0)$$
$$+ \ln \frac{\mathbb{P}_1\{\gamma_k = 0 | \psi_k\}}{\mathbb{P}_\infty\{\gamma_k = 0 | \psi_k\}}(1 - \gamma_k).$$

The adaptive censoring policy is given by

$$\psi_k = \begin{cases} \psi^*(1), & \text{if} \quad s_{k-1} \geq a_1 \\ \psi^*(\epsilon_1), & \text{if} \quad s_{k-1} < a_1 \end{cases} \qquad (9)$$

where $0 < \epsilon_1 \leq 1$ and $\psi^*(\epsilon_1)$ is defined as follows. Let $0 < e \leq 1$, define

$$\psi^*(e) = \arg\max_{\psi \in \mathcal{C}(e)} \mathbf{I}^\psi(f_1 || f_0) \qquad (10)$$

where $\mathcal{C}(\epsilon) = \{\psi : \mathbb{P}_\infty\{\psi(X_k) = 1\} = \epsilon\}$, and $\mathbf{I}^\psi(f_1 || f_0)$ is the K-L divergence of the observations available at the decision maker under the censoring strategy ψ

$$\mathbf{I}^\psi(f_1 || f_0) = \mathbb{E}_1\left[\ell^\psi(\gamma_1, X_1)\right].$$

Among the censoring strategies that have communication rate ϵ, the strategy $\psi^*(\epsilon)$ has the maximal post-censoring K-L divergence. In general, $\psi^*(\epsilon)$ does not have analytic expressions, but it is well known that $\psi^*(\epsilon)$ has a special structure: the likelihood ratio of the no-send region is *a single interval* [18]. The upper and lower bounds of this single interval is obtained via numerical simulations.

The stopping time for the CuSum-AC algorithm is given by

$$T_c(a) = \inf\{k : s_k \geq a\}. \qquad (11)$$

In the following, we use $T_c(x, y)$ to denote the stopping time as $T_c(a)$ for which the initial statistic is $s_0 = x$ and the threshold is y.

In summary, the CuSum-AC algorithm is illustrated in Fig. 3. A few remarks on the algorithm are presented as follows.

Compared with the CuSum algorithm, two additional parameters ϵ_1, a_1 are introduced for the CuSum-AC algorithm. As in the CuSum algorithm, the parameters involved in the CuSum-AC algorithm, i.e., ϵ_1, a_1 together with a, can only be determined by numerical simulations. In practice, if the required ARLFA is large enough (i.e., the threshold a is chosen large enough), the communication rate constraint and the
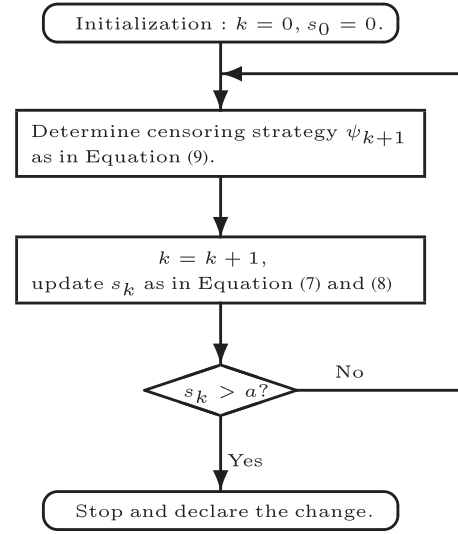

Fig. 3. CuSum-AC algorithm.

ARLFA constraint can be met independently. More specifically, to meet the communication rate constraint, one may first fix a large enough a (any value) and then obtain an appropriate pair (a_1, ϵ_1). Since given a pair (ϵ_1, a_1), when a is large enough, the communication rate is almost a constant, one then may meet the ARLFA constraint by just adjusting a.

The detection statistic of the CuSum-AC algorithm is reset to switching threshold a_1 whenever it crosses a_1 from below. This facilitates the asymptotic optimality analysis of the CuSum-AC algorithm and makes the stopping time $T_c(a)$ of the CuSum-AC algorithm an equalizer rule (the details of which are given in Section IV).

We now elaborate on the adaptive censoring strategy. Note that by definition, for a fixed $0 < e \leq 1$, $\psi^*(e)$ is the most informative in the sense that it achieves the largest post-censoring K-L divergence with communication rate e. While by adjusting e, one can trade off communication cost against information quality of $\psi^*(e)$. On the one hand, the larger e is, the more information of the observations taken at the sensor node is conveyed to the decision maker under $\psi^*(e)$. On the other hand, if $e_1 > e_2$, $\psi^*(e_1)$ incurs more communication cost than $\psi^*(e_2)$ does. Intuitively, one tends to use a censoring strategy $\psi^*(e)$ with a larger e when it is deemed "more important." At each time k, our adaptive censoring strategy in (9) tends to use a censoring strategy $\psi^*(e)$ with a larger e when s_{k-1} is larger. This idea comes from the observation of the typical evolution of the CuSum algorithm as illustrated in Fig. 4. The detection statistic c_k goes up and down before it reaches the threshold. At most times, the detection statistic stays small. Note that if the sojourn time when c_k stays in one interval is large enough, the change of c_k in that interval can be approximated using the *statistical property* of the observations without knowing each observation. Let us take two extreme cases for example. Let T_1 and T_2 be the sojourn time when c_k is in interval 1 and interval 2, respectively. Suppose that T_1 is sufficiently large, while $T_2 = 1$. Then by the renewal theorem [24], one knows that the change of c_k in interval 1 can be obtained by $\Delta_1(c_k) \approx T_1 \mathbf{I}(f_1 || f_0)$. This means that almost no information

---

[3]Note that since a censoring strategy is adopted, when $\gamma_k = 0$, the decision maker still has a rough information about X_k.
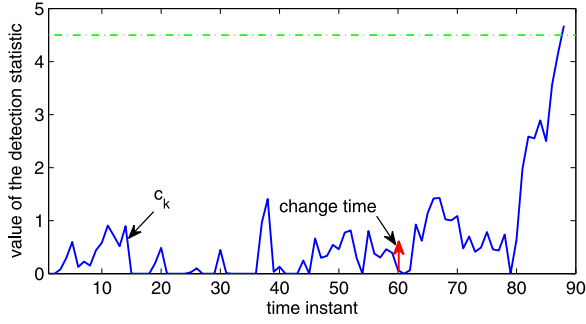
Fig. 4. Typical evolution of the CuSum algorithm. There is a mean shift in Gaussian noise, where the parameters used are as follows: before the change $\mathbb{P}_\infty : X_k \sim \mathcal{N}(0, 1)$, after the change $\mathbb{P}_1 : X_k \sim \mathcal{N}(0.5, 1)$, change time $\nu = 60$ and threshold $a = 4.5$.

is lost for the decision maker even if no messages are sent by the sensor. However, for the case $T_2 = 1$, in order to make the change $\Delta_2(c_k)$ known to the decision maker, this single observation has to be sent to the decision maker, the communication rate of which is 1. Based on this notion, one extends the adaptive censoring strategy to $N > 2$ levels as follows:

$$\psi_k = \begin{cases} \psi^*(1), & \text{if } s_{k-1} \geq a_1 \\ \psi^*(\epsilon_1), & \text{if } a_2 \leq s_{k-1} < a_1 \\ \vdots & \vdots \\ \psi^*(\epsilon_{N-2}), & \text{if } a_{N-1} \leq s_{k-1} < a_{N-2} \\ \psi^*(\epsilon_{N-1}), & \text{if } s_{k-1} < a_{N-1} \end{cases} \quad (12)$$

with $0 < \epsilon_{N-1} \leq \epsilon_{N-2} \leq \cdots \leq \epsilon_1 \leq 1$ and $a_{N-1} < a_{N-2} < \cdots < a_1$.

Remark 1: Now we discuss the practical implementation of the CuSum-AC algorithm with general N levels and the online computation load at the sensor side. The parameters, i.e., $\psi^*(\epsilon_{N-1}), \ldots, \psi^*(\epsilon_1), a_{N-1}, \ldots, a_1$ and a, are determined prior to the system run time. The censoring strategies $\psi^*(\epsilon_n)$, $1 \leq n \leq N - 1$ are stored in the sensor node[4] and the feedback message from the fusion center is the strategy index n (together with $n = 0$ representing $\psi^*(1)$). The feedback happens when $\psi_k \neq \psi_{k-1}$. Note that $\psi^*(\epsilon_n)$ has a special structure: the likelihood ratio of the no-send region is a single interval. Hence, to store a censoring strategy $\psi^*(\epsilon_n)$, it suffices to store the corresponding lower and upper bounds of the likelihood ratio (or the observations in some special cases, see below). The only computation task of the remote sensor is to implement the censoring strategies $\psi^*(\epsilon_n)$, the computational load of which is explained for the following two cases. For general distributions f_1 and f_0, the sensor first computes the likelihood ratio of X_k and then compares it to the upper and lower bounds. Note that to run the CuSum-like algorithms locally, the remote sensor needs to further compute the logarithm of the likelihood ratio. Since comparisons have negligible computational load compared with that of computing a logarithm, the computational load at the

---

[4]There is no need to store $\psi^*(1)$, under which no observations are censored at all.

remote senor for our algorithm is much lower. If the distributions f_1 and f_0 are such that the likelihood ratio function is monotone, then a single interval of the likelihood ratio also implies a single interval of the observations. Then to implement $\psi^*(\epsilon_n)$, the sensor just needs to compare the observations directly to the corresponding lower and upper bounds of the observations. The family of distributions that have monotone likelihood ratio property is quite large, e.g., exponential, Binomial, Poisson and normal distributions with known variances. Although, in most of these cases, the computational load of log-likelihood ratio is also low—one does not need to actually compute the logarithm but only elementary computations are required. Still, compared with running CuSum-like algorithms, the computational load at the remote sensor is considerably reduced in our algorithm, since comparisons are much simpler to compute than multiplications, especially when the observations take on real values.

IV. PERFORMANCE ANALYSIS

In this section, we first show that the $T_c(a)$ is an equalizer rule, i.e., the detection delay $d_L(T_c(a), \Psi)$ is attained for any change time ν. We then prove $T_c(a)$ is asymptotically optimal for any communication rate constraint.

A. Supporting Definitions

The classical performance analysis of the CuSum algorithm interprets the CuSum algorithm as a sequence of two-sided sequential probability ratio tests (SPRTs) [25]. This technique is also used for our analysis of the CuSum-AC algorithm. Intuitively, the CuSum-AC algorithm is a sequence of two-sided (0 and a) SPRTs with switching modes (original or censored) of observations. Each time the detection statistic crosses a_1 from below, it is reset to be a_1. This behavior is mathematically characterized as follows.

Define a stopping time of an SPRT with a starting point $0 \leq z < a - a_1$ as a variable

$$\eta(z) = \inf \left\{ n : z + \sum_{k=1}^n \ell(X_k) \notin [0, a - a_1] \right\}.$$

Note that $\eta(0)$ can be viewed as the first time that the detection statistic jumps out from $[a_1, a]$ with the initial point a_1. It either crosses the threshold a or returns to $[0, a_1]$ and starts a test with censored observations. We denote by $\hat{s}_{\eta(z)}$ the detection statistic at the time instant $\eta(z)$ bounded below by zero

$$\hat{s}_{\eta(z)} = \left(z + \sum_{k=1}^{\eta(z)} \ell(X_k) + a_1 \right)^+.$$

Define a detection statistic $\breve{s}_k(z)$, which is updated in the same manner with that in the CuSum algorithm but with an initial point $0 \leq z < a_1$ and censored observations. The details are as follows:

$$\breve{s}_k(z) = \left(\breve{s}_{k-1}(z) + \ell^{\psi^*(\epsilon_1)}(\gamma_k, X_k) \right)^+$$

$$\breve{s}_0(z) = z. \quad (13)$$

Based on $\breve{s}_k(z)$, we define a stopping time by

$$\phi(z) = \inf \left\{ k : \breve{s}_k(z) \geq a_1 \right\}.$$

As the CuSum-AC algorithm starts at 0, $\phi(0)$ can be interpreted as the first time that it reaches a_1 and switches the observation mode from the censored one to the original one.

Let

$$\Phi = \eta(0) + \phi\left(\hat{s}_{\eta(0)}\right) \mathbf{1}_{\left\{\hat{s}_{\eta(0)} < a_1\right\}}. \tag{14}$$

The CuSum-AC algorithm can be interpreted as a sequence of SPRTs with stopping times being of two distributions. Specifically, the CuSum-AC algorithm starts with the stopping time distributed as $\phi(0)$, and after the time instant $\phi(0)$, it is a sequence of SPRTs with stopping times i.i.d. distributed as Φ.

B. Equalizer Rule and Asymptotic Optimality

Theorem 1: The stopping time $T_c(a)$ is an equalizer rule for Problem 1, i.e.,

$$d_L\left(T_c(a), \Psi\right) = \operatorname*{ess\,sup}_{\mathcal{I}_{\nu-1}} \mathbb{E}_\nu\left[\left(T_c(a) - \nu + 1\right)^+ | \mathcal{I}_{\nu-1}\right] \quad \forall \nu \geq 1.$$

Remark 2: In general, the parameters used for the algorithm (i.e., ϵ_1, a_1 and a) can only be obtained by numerically simulating the detection performance (i.e., the delay and the ARLFA). The above theorem means that the change time does not affect the value of $d_L(T, \Psi)$. For simplicity, we can just let $\nu = 1$ to simulate the delay.

We now focus on asymptotic optimality of $T_c(a)$. Before presenting the main theorem, we first present the supporting lemma about the communication rate of the CuSum-AC algorithm as follows.

Lemma 1: Given any finite $a_1 > 0$ and $0 < \epsilon \leq 1$, there exists a nonempty set $\mathcal{E}(a_1, \epsilon)$ such that when $\psi^*(\epsilon_1)$ with $\epsilon_1 \in \mathcal{E}(a_1, \epsilon)$ is used, the communication rate constraint is uniformly satisfied for any $a > a_1$ (including $+\infty$). In other words, given any finite $a_1 > 0$ and $0 < \epsilon \leq 1$, there exists a censoring strategy Ψ as in (9) with $\epsilon_1 \in \mathcal{E}(a_1, \epsilon)$, such that

$$r(\Psi) \leq \epsilon \quad \forall a \in (a_1, +\infty].$$

The asymptotic optimality analysis involves the scenario where the threshold $a \to \infty$. The above theorem enables us to study the asymptotic performance of the CuSum-AC algorithm without worrying whether the communication constraint will be violated for some a.

Given a_1 and ϵ, we define a set $\mathcal{E}^*(a_1, \epsilon)$ as

$$\mathcal{E}^*(a_1, \epsilon) = \mathcal{E}(a_1, \epsilon) \cap \mathcal{E}'(a_1, \epsilon) \tag{15}$$

where the set $\mathcal{E}'(a_1, \epsilon)$ is given by

$$\mathcal{E}'(a_1, \epsilon) = \left\{ \epsilon_1 > 0 : \mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right) | \hat{s}_{\eta(0)} < a_1\right] \right.$$
$$\left. \geq \mathbb{E}_\infty\left[T(a_1)\right] \quad \forall a > a_1 \right\}.$$

Recall that $T(a_1)$ is the stopping time for the CuSum algorithm with a_1 as the threshold. Under Assumption 1, using the standard performance analysis technique for the CuSum algorithm (e.g., P.142 of [25]), one sees that $\mathbb{E}_\infty[T(a_1)]$ is finite for any finite a_1. Using the same analysis for $\mathcal{E}(a_1, \epsilon)$ (see the proof of Lemma 1), one sees that $\mathcal{E}'(a_1, \epsilon)$ is not empty for any a_1 and ϵ. Furthermore, when ϵ_1 is small enough, it must belong to both $\mathcal{E}(a_1, \epsilon)$ and $\mathcal{E}'(a_1, \epsilon)$, then $\mathcal{E}^*(a_1, \epsilon)$ is a non-empty set for any a_1 and ϵ.

We are now ready to present the second main theorem.

Theorem 2: For any $\epsilon > 0$, when $a = \ln \zeta$, $\epsilon_1 \in \mathcal{E}^*(a_1, \epsilon)$ with any $0 < a_1 < a$ are used, the CuSum-AC algorithm satisfies the ARFLA constraint (4) and communication rate constraint (5). Furthermore, the CuSum-AC algorithm is asymptotically ($\zeta \to \infty$) optimal for Problem 1, i.e., as $\zeta \to \infty$

$$d_L\left(T_c(\ln \zeta), \Psi\right) = \frac{\ln \zeta}{\mathbf{I}(f_1 \| f_0)} \left(1 + o(1)\right).$$

Remark 3: The globally (for any communication rate constraint $\epsilon > 0$) asymptotic optimality of the Cusum-AC algorithm stated in Theorem 2 relies on the following two factors: i) the asymmetric behavior of the detection statistic for the CuSum-AC algorithm s_k on pre-change and post-change hypotheses, ii) the fact that the communication rate defined in (2) is merely on the pre-change hypothesis. On the one hand, under $\mathbb{P}_\infty$, the expected duration of s_k being above a_1 is finite even when a is infinite (i.e., $T_{a_1}^\infty$ defined in (28) is finite), then one can choose ϵ_1 (equivalent to $\psi^*(\epsilon_1)$) to make the communication rate (ARLFA) of the CuSum-AC algorithm arbitrarily small (large). On the other hand, under $\mathbb{P}_1$, the expected duration of s_k being below a_1 is finite for any $\epsilon_1 > 0$, while the expected duration of s_k being above a_1 goes to infinity when $a \to \infty$.

Remark 4: Note that the general $N > 2$ levels cases provide more degrees of design freedom (including $a_{N-1}, \ldots, a_1$ and $\epsilon_{N-1}, \ldots, \epsilon_1$) than the two levels case (including a_1 and ϵ_1). Following exactly the same arguments for Theorem 2, one concludes that the CuSum-AC algorithm with $N > 2$ levels is asymptotically ($\zeta \to \infty$) optimal. Though the asymptotic optimality can be achieved for any censoring levels, better detection performance should be expected when N increases if ζ takes moderate values.

Remark 5: We remark that the feedback transmissions needed for the CuSum algorithm are in general quite few. In particular, the average number of feedback transmission under $\mathbb{P}_\infty$ is $\mathbb{E}_\infty[N_F] = 2/(1 - \mathbb{P}_\infty\{\hat{s}_{\eta(0)} < a_1\})$. From (32), the ratio of the average feedback transmission times to the ARLFA is $\mathbb{E}_\infty[N_F]/\mathbb{E}_\infty[T_c(a)] \leq 2/(\mathbb{E}_\infty[\eta(0)] + \mathbb{E}_\infty[\phi(\hat{s}_{\eta(0)})|\hat{s}_{\eta(0)} < a_1]\mathbb{P}_\infty\{\hat{s}_{\eta(0)} < a_1\})$, which is usually negligible since the term $\mathbb{E}_\infty[\phi(\hat{s}_{\eta(0)})|\hat{s}_{\eta(0)} < a_1]\mathbb{P}_\infty\{\hat{s}_{\eta(0)} < a_1\}$ is in general quite large.

V. EXTENSION TO SENSOR NETWORKS

In this section, first we modify the considered problem to sensor networks. Then we generalize the CuSum-AC algorithm presented in Section III and the results obtained in Section IV to this case.

A. Problem Formulation

The system is illustrated in Fig. 1. Let $\mathcal{M} = \{1, \ldots, M\}$. As in the one sensor case, it assumed that at sensor m, $\{X_{\{m,1\}}, \ldots, X_{\{m,\nu-1\}}\}$ are i.i.d. with pdf $f_{\{0,m\}}$ and $\{X_{\{m,\nu\}}, \ldots\}$ are i.i.d. with pdf $f_{\{1,m\}}$. As in [9] and [26], it is assumed that the change event affects all the sensors simultaneously at ν and the observations are independent across the sensors, conditioned on the change point.

Like γ_k in (1), let $\gamma_{\{m,k\}}$ be indicator whether or not the sensor m sends its observation $X_{\{m,k\}}$ to the fusion center. Let $\psi_{\{m,k\}}$ be the censoring strategy used at the sensor m at the time instant k, i.e.,

$$\gamma_{\{m,k\}} = \psi_{\{m,k\}}\left(X_{\{m,k\}}\right).$$

Let $\psi_k^{\mathcal{M}} = \{\psi_{\{1,k\}}, \ldots, \psi_{\{M,k\}}\}$ be the censoring strategies used at all the sensor nodes at time k and $\Psi^{\mathcal{M}}$ be the censoring policy along the horizon, i.e., $\Psi^{\mathcal{M}} = \{\psi_1^{\mathcal{M}}, \ldots, \psi_T^{\mathcal{M}}\}$.

The average communication rate before the change event happens for the network is defined by

$$r(\Psi^{\mathcal{M}}) = \limsup_{n \to \infty} \frac{1}{nM} \mathbb{E}_{\infty} \left[\sum_{k=1}^{n} \sum_{m=1}^{M} \gamma_{\{m,k\}} \mid T \geq n \right]. \quad (16)$$

In [14], the authors posed communication rate constraint for each channel in the multi-channel setting (the affected subset of the sensors is unknown). Since the change event affects all the sensors simultaneously in our case, we instead use the average communication rate of the whole network (16). Given T and $\Psi^{\mathcal{M}}$, the ARLFA is defined in the same way as in the one sensor case, i.e.,

$$g(T, \Psi^{\mathcal{M}}) = \mathbb{E}_{\infty}[T].$$

Let

$$\mathcal{I}_k^m = \left\{ \left(1, \gamma_{\{m,1\}}, \gamma_{\{m,1\}} X_{\{m,1\}}\right), \ldots, \right.$$
$$\left. \left(k, \gamma_{\{m,k\}}, \gamma_{\{m,k\}} X_{\{m,k\}}\right)\right\}$$

and $\mathcal{I}_k^{\mathcal{M}} = \{\mathcal{I}_k^1, \ldots, \mathcal{I}_k^M\}$. Then the Lorden's detection delay is defined by

$$d_L(T, \Psi^{\mathcal{M}}) = \sup_{1 \leq \nu < \infty} \left\{ \operatorname*{ess\,sup}_{\mathcal{I}_{\nu-1}^{\mathcal{M}}} \mathbb{E}_{\nu} \left[(T - \nu + 1)^+ \mid \mathcal{I}_{\nu-1}^{\mathcal{M}}\right] \right\}.$$

Then the problem we are interested in is as follow:

Problem 2:

$$\operatorname*{minimize}_{T, \Psi^{\mathcal{M}}} \quad d_L(T, \Psi^{\mathcal{M}})$$

$$\text{subject to} \quad g(T, \Psi^{\mathcal{M}}) \geq \zeta \quad (17)$$

$$r(\Psi^{\mathcal{M}}) \leq \epsilon \quad (18)$$

where $\zeta \geq 1$ is a given lower bound of the ARLFA.

B. CuSum-AC Algorithm for Multiple Sensors Case

We only present and focus on the CuSum-AC algorithm with $N = 2$ levels for sensor networks; $N > 2$ levels can be generalized as in the one sensor case. Let a and a_1 be two thresholds. The stopping time is computed as

$$T_c^{\mathcal{M}}(a) = \inf \left\{k : s_k^{\mathcal{M}} \geq a\right\} \quad (19)$$

where the detection statistic $s_k^{\mathcal{M}}$ is updated as follows:

$$\tilde{s}_k^{\mathcal{M}} = \max \left\{0, s_{k-1}^{\mathcal{M}} + \sum_{m=1}^{M} \ell^{\psi_{\{m,k\}}} \left(\gamma_{\{m,k\}}, X_{\{m,k\}}\right)\right\}$$

$$s_k^{\mathcal{M}} = \begin{cases} a_1, & \text{if} \quad s_{k-1}^{\mathcal{M}} < a_1 \text{ and } \tilde{s}_k^{\mathcal{M}} \geq a_1 \\ \tilde{s}_k^{\mathcal{M}}, & \text{otherwise} \end{cases}$$

$$s_0^{\mathcal{M}} = 0$$

and the censoring strategies are given by

$$\psi_{\{m,k\}} = \begin{cases} \psi^*(1), & \text{if} \quad s_{k-1}^{\mathcal{M}} \geq a_1 \\ \psi^*\left(\epsilon_{\{m,1\}}\right), & \text{otherwise} \end{cases} \quad (20)$$

with $0 < \epsilon_{\{m,1\}} \leq 1$. Note that the censoring strategies used at all the sensor nodes are switched simultaneously, which are adaptive to the detection statistic available at the fusion center. This helps reduce the times of feedback and the feedback message can be broadcasted to all the sensor nodes by the fusion center.

Let $X_k^{\mathcal{M}} = \{X_{\{1,k\}}, \ldots, X_{\{M,k\}}\}$ and $\epsilon_1^{\mathcal{M}} = \{\epsilon_{\{1,1\}}, \ldots, \epsilon_{\{M,1\}}\}$. Note that $X_k^{\mathcal{M}}$ and $\psi_k^{\mathcal{M}}$ can be regarded as the "vector version" of X_k and ψ_k in one sensor case, respectively. The CuSum-AC algorithm for the multiple sensors case is equivalent to its counterpart in one sensor case working with $X_k^{\mathcal{M}}$ and $\psi_k^{\mathcal{M}}$. Thus the following theorems are straightforward.

Theorem 3: The stopping time $T_c^{\mathcal{M}}(a)$ is an equalizer rule for Problem 2, i.e., for any $\nu \geq 1$

$$d_L\left(T_c^{\mathcal{M}}(a), \Psi^{\mathcal{M}}\right) = \operatorname*{ess\,sup} \mathbb{E}_{\nu} \left[\left(T_c^{\mathcal{M}}(a) - \nu + 1\right)^+ \mid \mathcal{I}_{\nu-1}^{\mathcal{M}}\right].$$

Define $\mathcal{E}_*^{\mathcal{M}}(\epsilon, a_1)$ as the counterpart of $\mathcal{E}^*(a_1, \epsilon)$. To see that $\mathcal{E}_*^{\mathcal{M}}(\epsilon)$ is not empty, one can pose an additional constraint that $\epsilon_{\{1,1\}} = \cdots = \epsilon_{\{M,1\}}$. Then the arguments follow straightforwardly from that of $\mathcal{E}^*(a_1, \epsilon)$.

Theorem 4: For any $\epsilon > 0$, when $a = \ln \zeta$, $\epsilon_1^{\mathcal{M}} \in \mathcal{E}_*^{\mathcal{M}}(\epsilon, a_1)$ with any $0 < a_1 < a$ are used, the CuSum-AC algorithm satisfies the ARFLA constraint (17) and communication rate constraint (18). Furthermore, the CuSum-AC algorithm is asymptotically optimal for Problem 2, i.e., as $\zeta \to \infty$

$$d_L\left(T_c^{\mathcal{M}}(\ln \zeta), \Psi^{\mathcal{M}}\right) = \frac{\ln \zeta}{\mathbf{I}\left(f_{\{1,m\}} \| f_{\{0,m\}}\right)} \left(1 + o(1)\right).$$

VI. NUMERICAL EXAMPLES

For simulations, we consider the problem of mean shift detection in Gaussian noise. It is assumed that $M = 3$ identical sensors are deployed and the pre-change and post-change distributions are $f_{\{0,m\}} \sim \mathcal{N}(0,1)$ and $f_{\{1,m\}} \sim \mathcal{N}(0.5, 1)$, respectively. For simplicity, in each example, the sensors use an identical censoring strategy.

Example 1: The asymptotic optimality of the CuSum-AC algorithm is examined. We compare the detection performance of the CuSum-AC algorithm with that of the CuSum algorithm
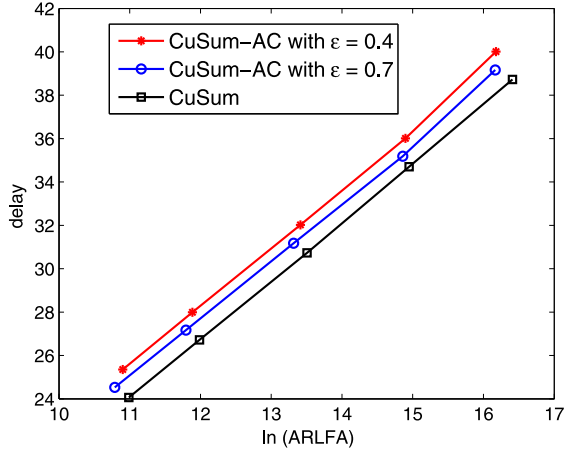
Fig. 5. Detection delay versus the ARLFA for the CuSum algorithm and the CuSum-AC algorithm.

Fig. 6. Detection delay versus the communication rate for the CuSum-AC algorithm and the CuSum algorithm with random transmission policy.

(which is the optimal one when there is no communication rate constraint). With different ARLFA's (sufficiently large), the detection delays (i.e., $\mathbb{E}_1[T]$) of these algorithms are simulated. For the CuSum-AC algorithm, two communication rate constraints are considered, i.e., $\epsilon = 0.7$ or $\epsilon = 0.4$. Note that given a communication rate and ARLFA constraint, there may exist multiple admissible combinations of the parameters (i.e., a, a_1, ϵ_1). To alleviate the computation burden, we set $\epsilon_1 = 0.63$ and $a_1 = 0.78$ for the case $\epsilon = 0.7$ and $\epsilon_1 = 0.27$ and $a_1 = 0.79$ for the case $\epsilon = 0.4$. The value of the threshold a is determined to make the communication rate constraint to be satisfied equally. Since given a_1 and ϵ_1, the communication rate is *not* strictly monotonic with a, multiple a's (which have different ARLFA's) can be found. In fact, given a_1 and ϵ_1, the communication rate remains the same when a varies if a is sufficiently large. The simulation results are given in Fig. 5. It can be seen that as the ARLFA increases, the difference between the delay of the CuSum-AC algorithm (with communication rate either $\epsilon = 0.7$ or $\epsilon = 0.4$) remains approximately constant. This verifies the asymptotic optimality, since the difference will be negligible when the ARLFA goes to infinity. Furthermore, we can see that for the CuSum-AC algorithm with the same ARLFA, the one which has the smaller communication rate (i.e., $\epsilon = 0.4$) has the larger detection delay. This is consistent with our intuition that better detection performance can be expected when more communication resources are used. We also note that when the communication rate is 0.4, the delay of the CuSum-AC algorithm is only around 1.2 time slots larger than that of the CuSum algorithm (the communication rate of which is 1). This shows good detection delay versus communication rate tradeoff for the CuSum-AC algorithm, which will be shown further in the next example.

Example 2: We compare the CuSum-AC algorithm with the CuSum algorithm with the random transmission policy. In the random transmission policy, whether an observation is transmitted or not is determined randomly such that the communication rate is satisfied. The random transmission policy is quite simple and serves as a lower bound of the detection performance in some sense. We plot Fig. 6 in the following way. The ARLFA is fixed to be 10 000, i.e., the parameters for the
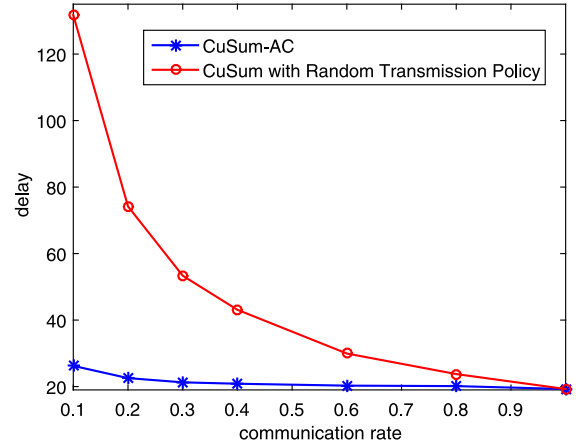
algorithm (a for the random transmission control policy, and a, a_1 and ϵ_1 for the CuSum-AC algorithm) should be chosen such that the associated ARLFA is around 10 000. The parameters for the CuSum-AC algorithm are determined using the brute-force search technique. Multiple admissible combinations of the parameters a, a_1 and ϵ_1 exist, among which the one that has the smallest detection delay is used. As depicted in Fig. 6, the CuSum-AC algorithm significantly outperforms the CuSum algorithm with the random transmission policy, in particular when the allowed communication rate is small. One also should note that the CuSum-AC algorithm has quite nice performance *per se*. In particular, when $\epsilon = 0.1$, the detection delay of the CuSum-AC algorithm is only 7 time slots larger than that of the CuSum algorithm (when there is no communication rate constraint, the CuSum-AC algorithm reduces to the CuSum algorithm).

VII. CONCLUSION AND FUTURE WORK

In this paper, we have studied the problem of minimax quickest change detection with communication rate constraints. The constraint is posed by limited energy at the remote sensors. An extension of the classical Lorden's formulation was studied. We proposed the CuSum-AC algorithm: the CuSum algorithm is used at the fusion center and adaptive censoring strategies are used at the sensor nodes. The CuSum-AC algorithm was proved to be an equalizer rule, and be globally asymptotically optimal for any positive communication rate constraint, as the ARLFA goes to infinity. The numerical simulations showed that the CuSum-AC algorithm has a better detection performance versus communication rate trade-off than the CuSum algorithm with random transmission control policy.

For the future work, there are two interesting directions. One is to explore the relationship between the detection performance when the ARLFA takes moderate values and the censoring strategy being used (in particular, to find whether there exists strictly optimal censoring strategy for every possible ARLFA). The other one is to study the problem in the multichannel setting (the change event only affects a subset of the sensors).

APPENDIX A
PROOF OF THEOREM 1

Before proceeding, we first give a supporting lemma as follows. Recall that the stopping time $T_c(z, a)$ can be viewed as the first time that the CuSum-AC algorithm reaches the threshold a, when starting at z. About $T_c(z, a)$, the follows hold.

Lemma 2: For any a and $0 \leq z < a$

$$\mathbb{E}_1 [T_c(z, a)] \leq \mathbb{E}_1 [T_c(a)]. \tag{21}$$

Proof: The proof is done by cases.

Case $0 \leq z < a_1$: Because of the reset action of the CuSum-AC algorithm when it crosses a_1 from below, for $0 \leq z < a_1$, we have

$$\mathbb{E}_1 [T_c(z, a)] = \mathbb{E}_1 [\phi(z)] + \mathbb{E}_1 [T_c(a_1, a)]. \tag{22}$$

The quantity $\mathbb{E}_1[T_c(a_1, a)]$ is a common term. Thus to obtain Lemma 2, we only need to prove that for any $0 \leq z < a_1$

$$\mathbb{E}_1 [\phi(z)] \leq \mathbb{E}_1 [\phi(0)] \tag{23}$$

which is easily obtained by sample path arguments.

Case $a_1 \leq z < a$: Starting at z, there are two possible ways for the CuSum-AC algorithm to reach the threshold a eventually. One is that the algorithm never returns to the censored region before it stops, i.e., $s_k \geq a_1$ along the whole horizon; we denote this event by $\mathcal{Z}_z^\uparrow$. The other one is that the detection statistic s_k once crosses a_1 from up before the algorithm stops, which is denoted by $\mathcal{Z}_z^\downarrow$. Let $p(z) = \mathbb{P}_1\{\mathcal{Z}_z^\uparrow\}$. We then have

$$
\begin{aligned}
&\mathbb{E}_1 [T_c(z, a)] \\
&= p(z)\mathbb{E}_1 [T_c(z, a)|\mathcal{Z}_z^\uparrow] + (1 - p(z)) \mathbb{E}_1 [T_c(z, a)|\mathcal{Z}_z^\downarrow] \\
&= p(z)\mathbb{E}_1 [T_c(z, a)|\mathcal{Z}_z^\uparrow] \\
&\quad + (1 - p(z)) \left(\mathbb{E}_1 [\eta(z)|\mathcal{Z}_z^\downarrow] \right. \\
&\qquad\qquad \left. + \mathbb{E}_1 [T_c(x, a); \hat{s}_{\eta(z)} = x|\mathcal{Z}_z^\downarrow] \right) \\
&= p(z)t^\uparrow + (1 - p(z)) \left(t_1^\downarrow + t_2^\downarrow \right).
\end{aligned}
$$

The physical meaning of $t_1^\downarrow$ is the conditional average time it takes for the CuSum-AC algorithm to cross a_1 from up for the first time, when starting at z.

Before proceeding, we first define a stopping time $\tilde{T}(a)$ as follows. This is a stopping time for a variant of the CuSum algorithm that works in the same manner with the CuSum algorithm but starts at a_1 and bounded below by a_1

$$\tilde{T}(a) = \inf\{k : \tilde{c}_k \geq a\}$$

where $\tilde{c}_k$ involves by

$$
\begin{aligned}
\tilde{c}_k &= \max(\tilde{c}_{k-1} + \ell(X_k), a_1) \\
\tilde{c}_0 &= a_1.
\end{aligned}
$$

For the quantities $p(z)$, $t^\uparrow$ and $t_1^\downarrow$, from Lemma 5 of [13], we then have an inequality: for any $t^* \geq \mathbb{E}_1[\tilde{T}(a)]$

$$p(z)t^\uparrow + (1 - p(z)) \left(t_1^\downarrow + t^* \right) \leq t^*. \tag{24}$$

The reset action for the CuSum-AC algorithm when crossing a_1 from below together with the fact that $\{X_k\}'s$ are i.i.d. under the measure $\mathbb{P}_1$ yields

$$\mathbb{E}_1 [T_c(a_1, a)] \geq \mathbb{E}_1 \left[\tilde{T}(a) \right].$$

Combining (22), one obtains

$$\mathbb{E}_1 [T_c(a)] \geq \mathbb{E}_1 \left[\tilde{T}(a) \right]. \tag{25}$$

We then study the quantity $t_2^\downarrow$

$$
\begin{aligned}
t_2^\downarrow &= \mathbb{E}_1 \left[T_c(x, a); \hat{s}_{\eta(z)} = x|\mathcal{Z}_z^\downarrow \right] \\
&= \int_{0 \leq x < a_1} \mathbb{E}_1 \left[T_c(x, a)|\mathcal{Z}_z^\downarrow; \hat{s}_{\eta(z)} = x \right] d\mathbb{P}_1 \left\{ \hat{s}_{\eta(z)} \leq x|\mathcal{Z}_z^\downarrow \right\} \\
&\stackrel{(e1)}{=} \int_{0 \leq x < a_1} \mathbb{E}_1 \left[T_c(x, a) \right] d\mathbb{P}_1 \left\{ \hat{s}_{\eta(z)} \leq x|\mathcal{Z}_z^\downarrow \right\} \\
&\stackrel{(ie1)}{\leq} \operatorname*{ess\,sup}_{0 \leq x < a_1} \mathbb{E}_1 \left[T_c(x, a) \right] \leq \mathbb{E}_1 [T_c(a)].
\end{aligned}
\tag{26}
$$

The equality $(e1)$ follows from the Markovity of the detection statistic for the CuSum-AC algorithm. Thus, given $\hat{s}_{\eta(z)} = x$, from the time instant $k = \eta(z)$ on, the evolution of the CuSum-AC algorithm, which starts at z, is exactly the same with that of a new CuSum-AC algorithm that starts at x. The inequality $(ie1)$ holds because of Hölder's inequality.

Then for any $a_1 \leq z < a$

$$
\begin{aligned}
\mathbb{E}_1 [T_c(z, a)] &= p(z)t^\uparrow + (1 - p(z)) \left(t_1^\downarrow + t_2^\downarrow \right) \\
&\stackrel{(ie1)}{\leq} p(z)t^\uparrow + (1 - p(z)) \left(t_1^\downarrow + \mathbb{E}_1 [T_c(a)] \right) \stackrel{(ie2)}{\leq} \mathbb{E}_1 [T_c(a)]
\end{aligned}
$$

where the inequality $(ie1)$ follows from (26) and $(ie2)$ follows from (24) and (34). The proof thus is completed. ∎

We are ready to prove Theorem 1. From the Markovity of the detection statistic s_k for the CuSum-AC algorithm, one can see that the average detection delay is measurable with respect to $s_{\nu-1}$, i.e.,

$$
\begin{aligned}
\mathbb{E}_\nu \left[(T_c(a) - \nu + 1)^+ |\mathcal{I}_{\nu-1} \right] &= \mathbb{E}_\nu \left[(T_c(a) - \nu + 1)^+ |s_{\nu-1} \right] \\
&= \mathbb{E}_\nu \left[(T_c(s_{\nu-1}, a)] \right.
\end{aligned}
$$

Note that under Assumption 1, for any censoring strategy $\psi^*(\epsilon_1)$, from the positive definiteness of K-L divergence [27], one has $\mathbb{E}_\infty[\ell^{\psi^*(\epsilon_1)}(\gamma_k, X_k)] < 0$. It then follows that:

$$\mathbb{P}_\infty \left\{ \ell^{\psi^*(\epsilon_1)}(\gamma_k, X_k) < 0 \right\} := p > 0.$$

and for any $\nu \geq 1$

$$\mathbb{P}_\infty\{s_{\nu-1} = 0\} \geq p^{\nu-1} > 0. \tag{27}$$

We then obtain

$$
\begin{aligned}
\operatorname*{ess\,sup}_{\mathcal{I}_{\nu-1}} &\mathbb{E}_\nu \left[(T_c(a) - \nu + 1)^+ |\mathcal{I}_{\nu-1} \right] \\
&= \operatorname*{ess\,sup}_{s_{\nu-1}} \mathbb{E}_\nu [T_c(s_{\nu-1}, a)] \\
&= \mathbb{E}_\nu [T_c(s_{\nu-1} = 0, a)]
\end{aligned}
$$

where the last equality follows from Lemma 2 and (27). The proof thus is completed.

APPENDIX B
PROOF OF LEMMA 1

We first give the upper bound of $r(\Psi)$ when a, a_1 and $\psi^*(\epsilon_1)$ are known. We then show that by choosing certain ϵ_1 (equivalently with $\psi^*(\epsilon_1)$), which is independent of a, the upper bound can be any admissible value. The technique used is to interpret the CuSum-AC algorithm as a sequence of two-sided SPRTs as in Section IV-A.

Define a random sequence $\{T_i\}$ as

$$T_0 = \phi(0)$$
$$T_i = T_{i-1} + W_i \quad \forall i \geq 1$$

where $W_i := W_i^{(1)} + W_i^{(2)}$. The random variables $W_i^{(1)}$ and $W_i^{(2)}$ are i.i.d. distributed with mean equal to $\mathbb{E}_\infty[\eta(0)|\hat{s}_{\eta(0)} < a_1]$ and $\mathbb{E}_\infty[\phi(\hat{s}_{\eta(0)})|\hat{s}_{\eta(0)} < a_1]$, respectively. Note that the distribution of W_i is *different* from that of Φ defined in (14). The stopping time Φ is for the evolution of the CuSum-AC algorithm (which may stop at some time, i.e., $\hat{s}_{\eta(0)} \geq a$), while for the definition of the communication rate constraint (2), we implicitly assume that the CuSum-AC algorithm never stops.

Based on $\{T_i\}$, we define a reward sequence $\{R_i\}$ as

$$R_0 = 0$$
$$R_i = R_{i-1} + W_i^{(1)} + J_i \quad \forall i \geq 1$$

where $J_i \sim B(W_i^{(2)}, \epsilon_1)$ is a binomial distributed random variable.

Given ϵ_1, a_1 and a

$$r(\Psi) \overset{(ie1)}{\leq} \limsup_{n\to\infty} \frac{1}{n - T_0} \mathbb{E}_\infty \left[\sum_{k=T_0+1}^{n} \gamma_k \middle| T \geq n \right]$$
$$:= \lim_{i\to\infty} \frac{R_i}{T_i - T_0} \overset{(e1)}{=} \frac{\mathbb{E}_\infty\left[W_1^{(1)} + J_1\right]}{\mathbb{E}_\infty[W_1]}$$
$$:= \frac{\mathbb{E}_\infty\left[\eta(0)|\hat{s}_{\eta(0)} < a_1\right] + \epsilon_1 \mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]}{\mathbb{E}_\infty\left[\eta(0)|\hat{s}_{\eta(0)} < a_1\right] + \mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]}$$
$$:= \bar{r}(\Psi)$$

where the inequality $(ie1)$ holds because by the definition of $\phi(0)$, before the time instant T_0, all the observations are censored using the censoring strategy $\psi^*(\epsilon_1)$, the communication rate of which is ϵ_1 (the lower bound of $r(\Psi)$); the equality $(e1)$ follows from the alternating renewal process theory ([28, p. 173]).

Using the sample path arguments, one sees that, given a_1, $\mathbb{E}_\infty[\eta(0)|\hat{s}_{\eta(0)} < a_1]$ is monotonically nondecreasing with a. Let

$$T_{a_1}^\infty := \lim_{a\to+\infty} \mathbb{E}_\infty\left[\eta(0)|\hat{s}_{\eta(0)} < a_1\right]. \quad (28)$$

Since $\mathbb{E}_\infty[\ell(X_k)] < 0$ (by Assumption 1), [24, Corollary 2.4] yields that $T_{a_1}^\infty$ is finite. Note that $\bar{r}(\Psi)$ is monotonically nondecreasing with $\mathbb{E}_\infty[\eta(0)|\hat{s}_{\eta(0)} < a_1]$, then for any a

$$\bar{r}(\Psi) \leq \frac{T_{a_1}^\infty + \epsilon_1 \mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]}{T_{a_1}^\infty + \mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]}$$
$$= \frac{\frac{T_{a_1}^\infty}{\mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]} + \epsilon_1}{\frac{T_{a_1}^\infty}{\mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]} + 1}$$
$$:= \bar{\bar{r}}(\Psi).$$

For any censoring strategy $\psi^*(\epsilon_1)$, $\mathbb{E}_\infty[\ell^{\psi^*(\epsilon_1)}(\gamma_k, X_k)] \leq 0$. Furthermore, as $\epsilon_1 \to 0$, $\mathbb{E}_\infty[\ell^{\psi^*(\epsilon_1)}(\gamma_k, X_k)] \to 0$.[5] Note that when $\epsilon_1 = 0$, not only the mean of the log-likelihood ratio but also the random variable $\gamma_k X_k$ reduces to constant 0. By [24, Corollary 2.6], one sees that as $\mathbb{E}_\infty[\ell^{\psi^*(\epsilon_1)}(\gamma_k, X_k)] \to 0$, $\mathbb{E}_\infty[\phi(z)] \to \infty$ for any $0 \leq z < a_1$. Note also that because of the reset action whenever the CuSum-AC algorithm crosses a_1 from below, $T_{a_1}^\infty$ is only related with the distribution of X_k under $\mathbb{P}_\infty$, which is independent of the censoring strategy $\psi^*(\epsilon_1)$ (i.e., ϵ_1). Given any ϵ, one thus can find a non-empty set $\mathcal{E}(a_1, \epsilon)$ that is independent of a (i.e., independent of the distribution of $\{\hat{s}_{\eta(0)}|\hat{s}_{\eta(0)} < a_1\}$) such that for any censoring strategy $\psi^*(\epsilon_1)$ with $\epsilon_1 \in \mathcal{E}(a_1, \epsilon)$, $\bar{r}(\Psi) \leq \epsilon$. The proof thus is completed.

APPENDIX C
PROOF OF THEOREM 2

By the definition of $\mathcal{E}^*(a_1, \epsilon)$, the CuSum algorithm satisfies the communication rate constraint (5). Note that the CuSum algorithm is strictly optimal for Problem 1 when $\epsilon = 1$. We prove Theorem 2 by relating the CuSum-AC algorithm to the CuSum algorithm.

Step 1: We first prove that the ARLFA of the CuSum-AC algorithm that uses the censoring strategy defined in Theorem 2 is *always larger* than that of the CuSum algorithm. This is due to the definition of $\mathcal{E}^*(a_1, \epsilon)$ in (15), by which ϵ_1 (equivalent to $\psi^*(\epsilon_1)$) is appropriately chosen. Note that the ARLFA can be made arbitrarily large by adjusting ϵ_1. To this end, we define a stopping time as follows:

$$\hat{T}(a) = \inf\{k : \hat{c}_k \geq a\}$$

where the detection statistic $\hat{c}_k$ evolves by

$$\hat{c}'_k = (\hat{c}_{k-1} + \ell(X_k))^+$$
$$\hat{c}_k = \begin{cases} a_1, & \text{if } \hat{c}'_k \geq a_1 \text{ and } \hat{c}_{k-1} < a_1 \\ 0, & \text{if } \hat{c}'_k < a_1 \text{ and } \hat{c}_{k-1} \geq a_1 \\ \hat{c}'_k, & \text{otherwise} \end{cases}$$
$$\hat{c}_0 = 0.$$

The difference between $\hat{c}_k$ and c_k in (6) for the CuSum algorithm is that $\hat{c}_k$ is reset to be a_1 (0) whenever it crosses a_1 from below (up). Using the sample path arguments, one can see that for any a

$$\mathbb{E}_\infty\left[\hat{T}(a)\right] \geq \mathbb{E}_\infty\left[T(a)\right]. \quad (29)$$

Let $N(k)$ be the number of $\hat{c}_k$ crossing a_1 from below by time instant k, i.e.,

$$N(k) = \begin{cases} N(k-1) + 1, & \text{if } \hat{c}_k \geq a_1 \text{ and } \hat{c}_{k-1} < a_1 \\ N(k-1), & \text{otherwise} \end{cases}$$

with $N(0) = 0$. Wald's identity [29] yields

$$\mathbb{E}_\infty\left[\hat{T}(a)\right] = (\mathbb{E}_\infty\left[\eta(0)\right]$$
$$+ T(a_1)\mathbb{P}_\infty\{\hat{s}_{\eta(0)} < a_1\})\mathbb{E}_\infty\left[N\left(\hat{T}(a)\right)\right]. \quad (30)$$

[5]This approach is not necessarily monotonic. In some cases, $\mathbb{E}_\infty[\ell^{\psi^*(\psi_1)}(\gamma_k, X_k)]$ can be zero for nonzero ϵ_1's.

Let $T_c^\diamond(a) = \inf\{k : s_k^\diamond \geq a\}$, where $s_k^\diamond$ is evolved in the same manner with s_k in (8) for the stopping time $T_c(a)$ except for the starting point. The detection statistic s_k starts at $s_0 = 0$, while $s_k^\diamond$ starts with a random variable $s_0^\diamond = x$ and the distribution of x is the same with that of the random variable $\{\eta(0)|\hat{s}_{\eta(0)} < a_1\}$. Using the sample path arguments, one sees that

$$\mathbb{E}_\infty[T_c(a)] \geq \mathbb{E}_\infty[T_c^\diamond(a)]. \tag{31}$$

Let $N^\diamond(k)$ be the number of $s_k^\diamond$ crossing a_1 from below by time instant k, which is defined in the same manner with $N(k)$. Also by Wald's identity, one has

$$\mathbb{E}_\infty[T_c^\diamond(a)]$$
$$= \left(\mathbb{E}_\infty[\eta(0)] + \mathbb{E}_\infty\left[\phi\left(\hat{s}_{\eta(0)}\right)|\hat{s}_{\eta(0)} < a_1\right]\mathbb{P}_\infty\{\hat{s}_{\eta(0)} < a_1\}\right)$$
$$\times \mathbb{E}_\infty[N^\diamond(T_c^\diamond(a))]. \tag{32}$$

Both $\hat{c}_k$ and $s_k^\diamond$ are reset to be a_1 when they cross a_1 from below, we thus obtain the following:

$$\mathbb{E}_\infty\left[N\left(\hat{T}(a)\right)\right] = \mathbb{E}_\infty[N^\diamond(T_c^\diamond(a))]$$
$$= \frac{1}{1 - \mathbb{P}_\infty\{\hat{s}_{\eta(0)} < a_1\}} \tag{33}$$

where the last equality follows because both $N(k)$ and $N^\diamond(k)$ are geometrically distributed.

Combining (15) and (29)–(33), one can see that for any ϵ, when the censoring strategy $\psi^*(\epsilon_1)$ with $\epsilon_1 \in \mathcal{E}^*(a_1, \epsilon)$ is used

$$\mathbb{E}_\infty[T_c(a)] \geq \mathbb{E}_\infty[T(a)]. \tag{34}$$

Then by the established performance results of the CuSum algorithm [25], the CuSum-AC algorithm satisfies the ARLFA constraint (4).

Step 2: We show that for any ϵ, as $a \to \infty$

$$\mathbb{E}_1[T_c(a)] = \mathbb{E}_1[T(a)](1 + o(1)).$$

The intuition is as follows. Since for any finite a_1 and $\epsilon_1 > 0$, the duration of s_k staying below a_1 (when censored observations are used) is finite, then as $a \to \infty$, the duration of s_k being above s_k (when raw observations) dominates. The asymptotic first-order behavior of the detection delay of the CuSum-AC algorithms thus resembles that of classical CuSum algorithm. Let

$$T_1 = \sum_{k=1}^{T_c(a)} \mathbf{1}_{\{s_k \geq a_1\}}$$
$$T_2 = T_c(a) - T_1.$$

Recall that s_k is the detection statistic for the CuSum-AC algorithm. The quantity T_1 (T_2) can be viewed as the duration that s_k stays above (below) a_1. Following similar arguments in *Step 1*, one obtains that:

$$\mathbb{E}_1[T_1] = \mathbb{E}_1[\eta(0)]\frac{1}{1 - \mathbb{P}_1\{\hat{s}_{\eta(0)} < a_1\}}$$
$$\mathbb{E}_1[T_2] \leq \mathbb{E}_1[\phi(0)]\frac{1}{1 - \mathbb{P}_1\{\hat{s}_{\eta(0)} < a_1\}}.$$

Since $\mathbf{I}(f_1||f_0) > 0$, by [24, Corollary 2.4], one obtains that as $a \to \infty$, $\mathbb{E}_1[\eta(0)] \to \infty$. Note that for any $\epsilon_1 > 0$, $\mathbf{I}^{\psi^*(\epsilon_1)}(f_1||f_0) > 0$. From the established performance analysis technique for the CuSum algorithm (e.g., [25, p. 142]), one sees that $\mathbb{E}_1[\phi(0)] < \infty$. Note that $\mathbb{E}_1[\phi(0)]$ is only related to a_1 and $\psi^*(\epsilon_1)$. By the definition of $\mathcal{E}^*(a_1, \epsilon)$, $\psi^*(\epsilon_1)$ is independent of a. The following thus can be obtained:

$$\frac{\mathbb{E}_1[T_2]}{\mathbb{E}_1[T_1]} \leq \frac{\mathbb{E}_1[\phi(0)]}{\mathbb{E}_1[\eta(0)]} \to 0, \quad \text{as } a \to \infty.$$

Then $a \to \infty$

$$\mathbb{E}_1[T_c(a)] = \mathbb{E}_1[T_1]((1 + o(1))). \tag{35}$$

Because of the reset action when s_k crosses a_1 from below, the following holds:

$$\mathbb{E}_1[T_1] \leq \mathbb{E}_1[T(a - a_1)]$$
$$= \frac{\ln(a - a_1)}{\mathbf{I}(f_1||f_0)}(1 + o(1)), \quad \text{as } a \to \infty. \tag{36}$$

Note that as $a \to \infty$

$$\mathbb{E}_1[T(a)] = \frac{\ln(a)}{\mathbf{I}(f_1||f_0)}(1 + o(1)). \tag{37}$$

Combining (35)–(37), one obtains that for any ϵ, as $a \to \infty$

$$\mathbb{E}_1[T_c(a)] \leq \mathbb{E}_1[T(a)](1 + o(1))$$
$$\overset{(e1)}{=} \mathbb{E}_1[T(a)](1 + o(1)) \tag{38}$$

where $(e1)$ holds because given (34), $\mathbb{E}_1[T(a)]$ is the lower bound of $\mathbb{E}_1[T_c(a)]$ as $a \to \infty$.

Step 3: By Theorem 1

$$d_L(T_c(a), \Psi) = \mathbb{E}_1[T_c(a)].$$

Note that $T(a)$ is asymptotically optimal for Problem 1, as $\zeta \to \infty$ (i.e., $a \to \infty$), when $\epsilon = 1$. Combining (34) and (38), one obtains that the CuSum-AC algorithm is asymptotically optimal. The proof thus is complete.

References

[1] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. 1st ACM Int. Workshop Wireless Sensor Netw. Appl.*, *ACM*, 2002, pp. 88–97.

[2] T. L. Lai, "Sequential changepoint detection in quality control and dynamical systems," *J. Roy. Statist. Soc., Ser. B (Methodological)*, vol. 57, pp. 613–658, 1995.

[3] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," *IEEE Trans. Signal Proces.*, vol. 54, no. 9, pp. 3372–3382, 2006.

[4] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. GLOBECOM*, 2008, pp. 1–5.

[5] A. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[6] A. N. Shiryaev, *Optimal Stopping Rules*. New York, NY, USA: Springer, 2007.

[7] G. Lorden, "Procedures for reacting to a change in distribution," *Annal. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[8] M. Pollak, "Optimal detection of a change in distribution," *Annal. Statist.*, vol. 13, pp. 206–227, 1985.

[9] V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[10] Y. Mei, "Quickest detection in censoring sensor networks," in *Proc. IEEE ISIT*, 2011, pp. 2148–2152.

[11] J. Geng and L. Lai, "Non-bayesian quickest change detection with stochastic sample right constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 5090–5102, Oct. 2013.

[12] T. Banerjee, V. Sharma, V. Kavitha, and A. Jayaprakasam, "Generalized analysis of a distributed energy efficient algorithm for change detection," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 91–101, 2011.

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

[14] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3727–3735, 2015.

[15] Y. Yilmaz, G. V. Moustakides, and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4509–4524, 2012.

[16] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[17] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*.   New York, NY, USA: Wiley, 2010.

[18] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, 1996.

[19] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in distributed and multi-channel systems," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal (ICASSP)*, 2013, pp. 3952–3956.

[20] M. Huang and S. Dey, "Dynamic quantizer design for hidden markov state estimation via multiple sensors with fusion center feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2887–2896, 2006.

[21] A. G. Tartakovsky, M. Pollak, and A. S. Polunchenko, "Third-order asymptotic optimality of the generalized Shiryaev-Roberts changepoint detection procedures," *Theory Probability Appl.*, vol. 56, no. 3, pp. 457–484, 2012.

[22] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[23] Y. Ritov, "Decision theoretic optimality of the Cusum procedure," *Ann. Statist.*, vol. 18, pp. 1464–1469, 1990.

[24] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia, PA: SIAM, 1982, vol. 39.

[25] H. V. Poor and O. Hadjiliadis, *Quickest Detection*.   Cambridge, U.K.: Cambridge Univ. Press, 2009, vol. 40.

[26] A. G. Tartakovsky and V. V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequent. Anal.*, vol. 27, no. 4, pp. 441–475, 2008.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.

[28] S. Asmussen, *Applied Probability and Queues*.   New York, NY, USA: Springer, 2003, vol. 51.

[29] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. New York, NY, USA: Springer, 1985.

Karl H. Johansson (F'13) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund, Sweden.

He is the Director of the Stockholm Strategic Research Area ICT The Next Generation and a Professor with the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at the University of California at Berkeley, California Institute of Technology, Nanyang Technological University, the Institute of Advanced Studies, Hong Kong University of Science and Technology, and the Norwegian University of Science and Technology. He has held a Senior Researcher Position with the Swedish Research Council. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation systems.

Dr. Johansson 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 the 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 the 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 was awarded Future Research Leader from the Swedish Foundation for Strategic Research in 2005, the Young Researcher Awards from Scania in 1996 and from Ericsson in 1998 and 1999, the triennial Young Author Prize from IFAC in 1996, and the Peccei Award from the International Institute of System Analysis, Austria, in 1993.

Dawei Shi received the B.Eng. degree in electrical engineering and automation from the Beijing Institute of Technology, Beijing, China, in 2008, and the Ph.D. degree in control systems from the University of Alberta, in 2014.

Since December 2014, he has been an Associate Professor at the School of Automation, Beijing Institute of Technology. His research interests include event-based control and estimation, robust model predictive control and tuning, and wireless sensor networks.

Dr. Shi is a reviewer for a number of international journals, including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica, and Systems & Control Letters*. In 2009, he received the Best Student Paper Award from the IEEE International Conference on Automation and Logistics.

Ling Shi (M'09) 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.

Dr. Shi was 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.

Xiaoqiang Ren received the B.E. degree from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2012. He is currently pursuing the Ph.D. degree in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology.

He was a visiting student in the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden, from January 2015 to June 2015, and then in the NeSC group, Zhejiang University, from August 2015 to January 2016. His research interests include sequential detection, security of cyber-physical systems, and n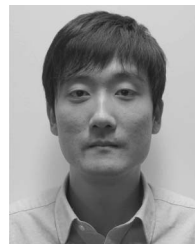etworked control and estimation.