Quickest Change Detection in Adaptive Censoring Sensor Networks

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

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 communication bandwidth, 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. Two minimax formulations are proposed. The cumulative sum (CuSum) algorithm is used at the fusion center and censoring strategies adaptive to the CuSum statistic are used at the sensor nodes. 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 for Lorden’s criterion and to be globally asymptotically optimal for any positive communication rate constraint for both formulations we propose, as the average run length to false alarm goes to infinity. It is also shown, by numerical examples, that the CuSum-AC algorithm has a good trade-off between the detection performance and the communication rate.

Keywords: censoring, quickest change detection, minimax, CuSum, asymptotically optimal, adaptive, wireless sensor networks.

*: Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong. Emails: xren, eesling@ust.hk.
†: ACCESS Linnaeus Center, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden. Email: kallej@ee.kth.se.
‡: State Key Laboratory of Intelligent Control and Decision of Complex Systems; School of Automation, Beijing Institute of Technology, Beijing, 100081, China. Email: daweishi@bit.edu.cn.
I. INTRODUCTION

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 condition monitoring [1], habitat monitoring [2], quality control engineering [3], computer security [4] and cognitive radio networks [5], [6]. 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 \( \nu \). 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 [7], [8] and the minimax formulation due to Lorden [9] and Pollak [10].

In the minimax formulation, \( \nu \) is an unknown constant and the false detection is formulated as the average run length to false alarm (ARLFA). With the constraint that the ARLFA is bounded below by \( \zeta \), Lorden’s problem [9] is to find the optimal stopping time \( T \) such that the worst-worst case detection delay \( \sup_{1 \leq \nu < \infty} \{ \esssup \mathbb{E}_\nu[(T - \nu + 1^+) | X_1, \ldots, X_{\nu-1}] \} \) is minimized, while Pollak’s problem [10] is to minimize the conditional worst case detection delay \( \sup_{1 \leq \nu < \infty} \mathbb{E}_\nu[(T - \nu + 1^+) | T \geq \nu] \). It is well known that the cumulative sum (CuSum) algorithm is strictly optimal, for any \( \zeta \geq 1 \), for Lorden’s problem [11], [12] and asymptotically optimal, as \( \zeta \to \infty \), for Pollak’s problem [13].

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 [14]–[16]. 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.

There are mainly two classes of characterizations of the cost of acquiring observations in the
literatures: the cost of sampling [16]–[20] and the cost of communication [21]–[23]. Premkumar and Kumar [17] formulated the cost as the expected accumulative number of sensors used in the sensor network before the system stops. Average number of observations taken before the change time for only one sensor node was considered by Banerjee and Veeravalli [18]. Geng et al. [19] studied a deterministic sampling constraint. Specifically, not only the expected value but also each realization of the sampling times was bounded from above. A criteria called pre-change duty cycle (the ratio of the active duration to the whole duration before change) was considered for the minimax settings in [20]. Geng and Lai [16] studied the scenario where the deployed sensor is replenished using energy harvesting technologies. Both [21] and [22] considered the energy constraints on analog transmissions for the sensor nodes. Banerjee and Veeravalli [23] considered both the cost of sampling and the cost of communication. The cost of sampling is formulated as in [20] and the cost of the communication is the ratio of transmission times to the whole time interval before the change. Note that [17], [21]–[23] studied the multiple sensors case, while one sensor or centralized cases were considered in [16], [18]–[20].

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 negligible compared with that of communication [24], [25]. 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. In the existing literature [21]–[23], to satisfy the communication rate constraint, the sensors only transmit when their local CuSum (or data efficient CuSum) statistics are above a certain threshold. The assumption is that the sensors have sufficient computation capacity such that these algorithms can run locally. This paper, however, focuses on the case where the remote sensors that have limited computation capacity and cannot run the CuSum algorithm, which involves computation of the likelihood ratio of the observations.

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 the limited communication bandwidth, the remote sensors cannot transmit at all the time instants. To make the best use of the limited resources, the sensors are assumed to adopt censoring strategies [26]. 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. For simplicity of presentation, first we consider the one sensor case. Then we generalize the results obtained to the multiple sensors scenario.

The main contributions of this work are summarized as follows.

- For quickest change detection with communication rate constraints, we propose the novel CuSum with adaptive censoring strategy (CuSum-AC) algorithm. For Lorden’s criterion, the algorithm is shown to be an equalizer rule (Theorem 1), i.e., the worst-worst case detection delay is attained for any change time. The detection delay is attained when the change event happens at $\nu = 1$ for Pollak’s criterion (Corollary 1). This property means that the CuSum-AC algorithm might be a good candidate for the optimal strategy. From the practical point of view, the computation work required to determine the optimal parameters of the algorithm can be reduced significantly. This is because that as in the CuSum algorithm, the parameters for the CuSum-AC algorithm can only be determined by numerically evaluating the associated performance metrics. By the aforementioned property, the detection delay for both problems can be computed by letting the change event happens at $\nu = 1$.

- We show that the CuSum-AC algorithm with carefully designed parameters is globally asymptotically (the ARLFA goes to infinity) optimal for any positive communication rate constraint (Theorem 2 and Theorem 3). A surprising point is that the asymptotically optimal censoring strategy may not necessarily fully utilize the available communication resources, i.e., the communication rate of the asymptotically optimal CuSum-AC algorithm is not necessarily equal to the upper bound of the communication rate constraint.

- We generalize all the above results obtained for the one sensor case to the sensor networks scenario. Specifically, the CuSum-AC algorithm for the multiple sensors case is an equalizer
rule for the Lorden’s criterion (Theorem 4) and globally asymptotically optimal for any positive communication constraint (Theorem 5).

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 for Lorden’s problem with communication constraint, 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.

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

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. Along the time horizon, a sequence of observations $\{X_k\}_{k \in \mathbb{N}_+}$ about the monitored environment are taken locally at the sensor. Let $\{X_k\}_{k \in \mathbb{N}_+}$ be a series of random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\nu$ is an unknown (but not random) time instant when a change event takes place. The instant $\nu$ may be $\infty$, corresponding to that the change never happens. The observations at the sensor before $\nu$, $\{X_1, \ldots, X_{\nu-1}\}$, are i.i.d. with measure $\mathbb{P}_\infty$, and the observations from $\nu$ on, $X_\nu, X_{\nu+1}, \ldots$, are i.i.d. with measure $\mathbb{P}_1$. Let $\mathbb{P}_\nu$ denote the probability measure when the change happens at $\nu$. 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. For simplicity, it is assumed that $\mathbb{P}_\infty$ and $\mathbb{P}_1$ are mutually locally absolutely continuous.

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 $\gamma_k$ as

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

Then the information pattern available for the decision maker at the time instant $k$ is given by $I_k = \{\gamma_1, \ldots, \gamma_k, \gamma_1X_1, \ldots, \gamma_kX_k\}$. A random variable $T \in \mathbb{N}_+$ is called a stopping time if $\{T = k\} \in \sigma(I_k)$, where $\sigma(I_k)$ is the smallest sigma-algebra of $I_k$. 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 $\theta_k$, is determined by the decision maker based on $I_{k-1}$. When $\theta_k \neq \theta_{k-1}$, the decision maker sends $\theta_k$ 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 $\theta_k : \Omega \mapsto \{0, 1\}$ has the form as $\gamma_k = \theta_k(X_k)$. The censoring policy along the horizon is given by $\Theta = \{\theta_1, \ldots, \theta_T\}$.

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

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

where $0 < \epsilon \leq 1$ is a design parameter. By adjusting $\epsilon$, a tradeoff between communication resources and detection performance is obtained. Note that the cost of the system still working after the change is penalized by the detection delay (see Problems 1 and 2 in the following), we
only pose the communication constraint before the change. The conditional expectation $\mathbb{E}_\infty [\cdot | \nu \geq n]$ thus is considered. Compared with sup, the communication rate constraint is less conservative when $\lim \sup$ is used. A similar criterion called pre-change transmission cost is considered in [27].

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 $\Theta$, the risk of false detection is characterized by the ARLFA

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

cf., [9], [28]. Note that the reciprocal of the ARLFA is connected to the false alarm rate. When $\zeta \to \infty$, the false alarm rate goes to zero. The stopping time $T$ is related to $I_k$, which is determined by the underlying observation sequence $\{X_k\}$ and the censoring policy $\Theta$, so the ARLFA is related with $\Theta$. To highlight this dependence, we use $g(T, \Theta)$ in the above definition.

The detection delay can be characterized by Lorden’s worst-worst case detection delay [9] and Pollak’s worst case conditional average delay [10]. The former is given by

$$d_L(T, \Theta) = \sup_{1 \leq \nu < \infty} \{ \esssup \mathbb{E}_\nu [(T - \nu + 1)^+ I_{\nu - 1}] \}, \quad (3)$$

and the latter by

$$d_P(T, \Theta) = \sup_{1 \leq \nu < \infty} \mathbb{E}_\nu [(T - \nu + 1)^+ | T \geq \nu]. \quad (4)$$

Since $\{T \geq \nu\} \in \sigma(I_{\nu-1})$ from the above definition of the stopping time, for any $T$ and $\Theta$, the following always holds: $d_P(T, \Theta) \leq d_L(T, \Theta)$. Using the above two characterizations of the detection delay, we have the following two problem formulations. The first one is the classical Lorden’s formulation with an additional communication rate constraint.

**Problem 1.**

$$\min_{T, \Theta} d_L(T, \Theta),$$

subject to

$$g(T, \Theta) \geq \zeta,$$

$$r(\Theta) \leq \epsilon,$$

where $\zeta \geq 1$ is a given lower bound of the ARLFA.
The second one is the classical Pollak’s formulation with an additional communication rate constraint.

**Problem 2.**

\[
\begin{align*}
\text{minimize} & \quad d_P(T, \Theta), \\
\text{subject to} & \quad g(T, \Theta) \geq \zeta, \\
& \quad r(\Theta) \leq \epsilon,
\end{align*}
\]

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

Note that since \( \Theta \) is adaptive, the available observation sequence \( \{\gamma_k X_k\} \) cease to be i.i.d. (the assumption for the classical formulation of the quickest change detection), which adds difficult to solving the above problems.

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

**Assumption 1.**

\[ 0 < I(\mathbb{P}_1 || \mathbb{P}_\infty) < \infty, \quad 0 < I(\mathbb{P}_\infty || \mathbb{P}_1) < \infty, \]

where

\[
\begin{align*}
I(\mathbb{P}_1 || \mathbb{P}_\infty) &= \int_\Omega \ln \frac{d\mathbb{P}_1}{d\mathbb{P}_\infty} d\mathbb{P}_1, \\
I(\mathbb{P}_\infty || \mathbb{P}_1) &= \int_\Omega \ln \frac{d\mathbb{P}_\infty}{d\mathbb{P}_1} d\mathbb{P}_\infty,
\end{align*}
\]

are the Kullback-Leibler (K-L) divergences. Both \( \frac{d\mathbb{P}_1}{d\mathbb{P}_\infty} \) and \( \frac{d\mathbb{P}_\infty}{d\mathbb{P}_1} \) in the above equations are the Radon-Nikodym derivative.

Our subsequent analysis utilizes the CuSum algorithm, which is stated in Fig. 3. The constant \( a \) is a given threshold and \( \ell(X_k) = \ln \frac{d\mathbb{P}_1}{d\mathbb{P}_\infty}(X_k) \) is the log-likelihood ratio function. The stopping time for the CuSum algorithm thus is computed as

\[
T_c(a) = \inf\{k : c_k > a\}. \tag{5}
\]

The CuSum algorithm is optimal when \( a \) is chosen such that \( \mathbb{E}_\infty[T_c(a)] = \zeta \). When there is no communication rate constraint, i.e., \( \epsilon = 1 \), Problems 1 and 2 are reduced to Lorden’s and
Pollak’s original formulations, respectively. We should remark that it is difficult to find strictly optimal algorithms for either Problem 1 or 2 (note that even the strictly optimal solution to Pollak’s problem is still an open problem). We hence focus on asymptotically optimal solution to the above two problems.

III. **CuSum Algorithm with an Adaptive Censoring Strategy**

In this section, we present the CuSum-AC algorithm. In the next section, we show that this CuSum-AC algorithm is asymptotically ($\zeta \to \infty$) optimal for any given $\epsilon$ for both Problems 1 and 2.

In [29], we studied the CuSum algorithm with stationary censoring strategy (CuSum-SC) and showed that given a communication rate constraint, the CuSum algorithm coupled with the censoring strategy that maximizes the post-censoring K-L divergence is asymptotically optimal (the asymptotical detection performance depends on the communication rate though; see Corollary 2), as the ARLFA goes to infinity, for both Lorden’s and Pollak’s problems. Since $\theta_k$ is fixed for the stationary censoring strategy, the post-censoring K-L divergence is given by

$$ I^\Theta(P_1 || P_\infty) = E_k \left[ \ell^{\theta_k}(\gamma_k, X_k) \right] $$

$$ = E_1 \left[ \ell^{\theta_1}(\gamma_1, X_1) \right] $$

$$ = I^{\theta_1}(P_1 || P_\infty), $$

where $\ell^{\theta_k}(\gamma_k, X_k)$ is the log-likelihood ratio function of the random variable $\gamma_k X_k$ under the
Stop and declare the change
\[ k = k + 1 ; \]
\[ s_k > a ? \]
No
Yes
Update \( s_k \) as in Equation (7)
Determine censoring strategy \( \theta_{k+1} \) as in Equation (8)
Initialization: \( k = 0 , s_0 = 0 \)

Fig. 4: CuSum-AC algorithm

censoring strategy \( \theta_k \) and is defined by
\[
\ell^{\theta_k}(\gamma_k, X_k) = \begin{cases} 
\ln \frac{d\mathbb{P}_1}{d\mathbb{P}_\infty}(X_k), & \text{if } \gamma_k = 1, \\
\ln \frac{\mathbb{P}_1(\gamma_k=0)}{\mathbb{P}_\infty(\gamma_k=0)}, & \text{if } \gamma_k = 0.
\end{cases}
\]

We now present the CuSum-AC algorithm. Let
\[
\theta^*(\epsilon) = \arg \max_{\theta \in C(\epsilon)} I^\theta(\mathbb{P}_1||\mathbb{P}_\infty),
\]
(6)

where \( C(\epsilon) = \{ \theta : \mathbb{P}_\infty \{ \theta(X_k) = 1 \} = \epsilon \} \). Among the censoring strategies that have communication rate \( \epsilon \), the strategy \( \theta^*(\epsilon) \) has the maximal post-censoring K-L divergence. Let \( \vec{a}_\downarrow = [a_1, \ldots, a_N] \) with \( 0 = a_N < a_{N-1} < \ldots < a_1 < \infty \) be a threshold vector. We define function \( \mathcal{F} : \mathbb{R}_+ \times \mathbb{R}^N_+ \mapsto \mathbb{R}_+ \) as
\[
\mathcal{F}(x, \vec{a}_\downarrow) = \max \{ a_i : x \geq a_i \}.
\]

Then the CuSum-AC algorithm with \( N > 1 \) levels is given in Fig. 4. The quantities \( a \) and \( \vec{a}_\downarrow \) are given thresholds with \( a_1 < a \). The detection statistic \( s_k \) are updated as follows:
\[
\tilde{s}_k = \max \{ 0, s_{k-1} + \ell^{\theta_k}(\gamma_k, X_k) \}, \\
s_k = \begin{cases} 
\mathcal{F}(\tilde{s}_k, \vec{a}_\downarrow), & \text{if } \mathcal{F}(\tilde{s}_k, \vec{a}_\downarrow) > \mathcal{F}(s_{k-1}, \vec{a}_\downarrow), \\
\tilde{s}_k, & \text{otherwise},
\end{cases}
\]
(7)
and the censoring strategy is given by

$$\theta_k = \begin{cases} 
\theta^*(1), & \text{if } s_{k-1} \geq a_1, \\
\theta^*(\epsilon^1), & \text{if } a_2 \leq s_{k-1} < a_1, \\
& \vdots \\
\theta^*(\epsilon_{N-2}), & \text{if } a_{N-1} \leq s_{k-1} < a_{N-2}, \\
\theta^*(\epsilon_{N-1}), & \text{if } s_{k-1} < a_{N-1},
\end{cases} \quad (8)$$

with $0 < \epsilon_{N-1} \leq \epsilon_{N-2} \leq \cdots \leq \epsilon_1 \leq 1$. The stopping time for the CuSum-AC algorithm thus is given by

$$T_{c-t} = \inf \{ k : s_k \geq a \}.$$  

**Remark 1.** The censoring strategy $\theta^*(\epsilon_n), 2 \leq n \leq N-1$ has communication rate $\epsilon_n$. Intuitively, better performance can be expected when more communication resources are used, i.e., the censoring strategy $\theta^*(\epsilon_n)$ conveys more information of the observations taken at the sensor node to the decision maker than $\theta^*(\epsilon_{n-1})$ does. In the CuSum-AC algorithm, the larger the detection statistic $s_k$ is, the more communication resources are spent. In particular, when the detection statistic is large enough ($s_{k-1} \geq a_1$), no censoring strategy is used and the original observations are directly transmitted.

**Remark 2.** The aforementioned idea of the censoring strategy for the CuSum-AC algorithm comes from the observation of the typical evolution of the CuSum algorithm as illustrated in Fig. 5. 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. Suppose that the sojourn time when $c_k$ is in interval 1, $T_1$, is sufficiently large and the sojourn time when $c_k$ is in interval 2, $T_2$, equals 1. Then by renewal theorem, one knows that the change of $c_k$ in interval 1 can be obtained by $\Delta_1(c_k) \approx T_1 I(\mathbb{P}_1||\mathbb{P}_\infty)$. This means that almost no information 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.
Fig. 5: 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 $a = 4.5$.

**Remark 3.** Generally, the sojourn time when the detection statistic is smaller is longer than that when the detection statistic is larger. Intuitively, the longer sojourn time enables the original observations to be approximated with “less accurate” observations. The communication rate $\epsilon_i$ thus is set to be monotonically increasing with the detection statistic for the censoring strategy defined in (8) to comply with this intuition.

**Remark 4.** The detection statistic is reset to the nearest threshold whenever it crosses that threshold from below. This facilitates the asymptotic optimality analysis of the CuSum-AC algorithm. What is more interesting, this reset action makes the stopping time $T_{c-\epsilon}$ of the CuSum-AC algorithm with $N = 2$ levels be an equalizer rule (the details of which is given in Section IV), which can facilitate the practical evaluation (using Monte Carlo experiments) of the involved performance metrics.

### IV. Main Results

For brevity of presentation, first we show the properties of the CuSum-AC algorithm with $N = 2$ levels. We show that the stopping rule is an equalizer rule for Problem 1 and asymptotically ($\zeta \to \infty$) optimal for any given $\epsilon$ for both Problems 1 and 2. We then extend the results to $N > 2$ levels.

To make (7) more clear for $N = 2$ levels, we rewrite it as follows:

$$s_k = \begin{cases} a_1, & \text{if } s_{k-1} < a_1 \text{ and } \tilde{s}_k \geq a_1, \\ \tilde{s}_k, & \text{otherwise,} \end{cases}$$
A. Equalizer Rule

In the following theorem, we prove that the \( T_{c-t} \) is an equalizer rule, i.e., the detection delay \( d_L(T, \Theta) \) is attained for any change time \( \nu \). As a corollary, we show that \( d_P(T, \Theta) \) is attained when \( \nu = 1 \).

The classical performance analysis of the CuSum algorithm interprets the CuSum algorithm as a sequence of two-sided sequential probability ratio tests (SPRTs) [30], so does the performance analysis of the DE-CuSum algorithm [20]. This technique is also used for our analysis of the CuSum-AC algorithm. Intuitively, the CuSum-AC algorithm is a sequence of two-sided \((0 \text{ and } a)\) 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 \( \hat{s}_k(z) \), 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:

\[
\hat{s}_k(z) = \left(\hat{s}_{k-1}(z) + \ell^{\theta_1}(\gamma_k, X_k)\right)^+,
\]

\[
\hat{s}_0(z) = z.
\]

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

\[
\phi(z) = \inf \{ k : \hat{s}_k(z) \geq a_1 \}.
\]

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

\[ \Phi = \eta(0) + \phi(\hat{s}_n(0)) \mathbb{1}_{\{\hat{s}_n(0) < a_1\}}. \]  

(10)

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 \( \Phi \).

Let

\[ T_{c-t}(z) = \inf\{k : s_k(z) \geq a\}, \]

where \( 0 \leq z < a \) and \( s_k(z) \) is updated in the same manner with \( s_k \) (the detection statistic for the stopping time \( T_{c-t} \) when \( N = 2 \)) but with the starting point \( s_0(z) = z \) instead. The stopping time \( T_{c-t}(z) \) can be viewed as the first time that the CuSum-AC algorithm reaches the threshold \( a \), when starting at \( z \). About \( T_{c-t}(z) \), we have the following lemma.

**Lemma 1.** For any \( 0 \leq z < a \),

\[ \mathbb{E}_1[T_{c-t}(z)] \leq \mathbb{E}_1[T_{c-t}]. \]  

(11)

**Proof:** See Appendix A. ■

**Remark 5.** The above lemma tells that the expectation under \( \mathbb{P}_1 \) of any CuSum-AC algorithm with non-zero starting point is bounded above by that of the CuSum algorithm that starts at 0.

We are ready to present the first main theorem.

**Theorem 1.** The stopping time \( T_{c-t} \) is an equalizer rule for Problem \([1]\) i.e.,

\[ d_L(T_{c-t}, \Theta) = \text{ess sup} \mathbb{E}_\nu[(T_{c-t} - \nu + 1)^+ | \mathcal{I}_{\nu - 1}], \quad \forall \nu \geq 1. \]

**Proof:** See Appendix B. ■

**Remark 6.** In general, the detection delay \( d_L(T, \Theta) \) can only be computed by numerical simulations. The above theorem means that the change time does not affect the value of \( d_L(T, \Theta) \). For simplicity of simulations, we can just let \( \nu = 1 \).
As a corollary, we show that the detection delay for Problem 2 is attained when the change event happens at \( \nu = 1 \).

**Corollary 1.** For the stopping time \( T_{c-t} \), the following holds:

\[
d_P(T_{c-t}, \Theta) = \mathbb{E}_1[T_{c-t}].
\]

**Proof:** For any \( \nu \geq 1 \)

\[
\mathbb{E}_\nu[(T_{c-t} - \nu + 1)^+|T_{c-t} \geq \nu] \leq \text{ess sup} \mathbb{E}_\nu[(T_{c-t} - \nu + 1)^+|I_{\nu-1}]
\]

\[
= \mathbb{E}_\nu[T_{c-t}(s_{\nu-1} = 0)]
\]

\[
= \mathbb{E}_1[T_{c-t}].
\]

The proof thus is complete.

\[\blacksquare\]

### B. Asymptotic Optimality

In this subsection, we first prove that the CuSum-AC algorithm with \( N = 2 \) levels is asymptotically (\( \zeta \to \infty \)) optimal for any given communication constraint \( \epsilon \). We then extend it to the CuSum-AC algorithm with \( N > 2 \) levels, i.e., the CuSum-AC algorithm with \( N > 2 \) levels is also asymptotically optimal. In the end, we analytically compare the asymptotic performance of the CuSum-AC algorithm with other two algorithms: one is the CuSum-SC algorithm proposed in our previous work [29] and the other one is the CuSum algorithm with random transmission control policy.

Before presenting the main theorem, we first give the supporting lemma and proposition as follows.

**Proposition 1.** Given any finite \( a_1 > 0 \) and \( 0 < \epsilon \leq 1 \), \( \exists\ \theta^*(\epsilon_1) \) with \( 0 < \epsilon_1 \leq \epsilon \) for the censoring strategy \( \Theta \) of the CuSum-AC algorithm with \( N = 2 \) levels such that the communication rate constraint is uniformly satisfied for any \( a > a_1 \) (including \( +\infty \)). In other words, given any finite \( a_1 > 0 \) and \( 0 < \epsilon \leq 1 \), \( \exists\ \Theta \), such that

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

**Proof:** See Appendix C.

\[\blacksquare\]
**Remark 7.** 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 $a$.

**Lemma 2.** For any censoring strategy $\theta^*(\epsilon_1)$ with $0 < \epsilon_1 \leq 1$,

$$
\epsilon_1 I(P_1 || P_\infty) \leq I^{\theta^*(\epsilon_1)}(P_1 || P_\infty) \leq I(P_1 || P_\infty).
$$

**Proof:** $I^{\theta^*(\epsilon_1)}(P_1 || P_\infty) \leq I(P_1 || P_\infty)$ follows from the invariance properties of K-L divergence \[31\].

We prove $\epsilon_1 I(P_1 || P_\infty) \leq I^{\theta^*(\epsilon_1)}(P_1 || P_\infty)$ by constructing the random censoring strategy $\Theta_r(\epsilon_1)$. The censoring strategy $\Theta_r(\epsilon_1)$ is stationary and is given by

$$
\gamma_k = \begin{cases} 
1, & \text{if } \xi \leq \epsilon_1, \\
0, & \text{otherwise},
\end{cases}
$$

(12)

where $\xi \sim unif(0, 1)$ is uniformly distributed and is independent of $X_k$. From the definition of $\theta^*(\epsilon_1)$, one can obtain that

$$
I^{\theta^*(\epsilon_1)}(P_1 || P_\infty) = \max_{\theta \in C(\epsilon_1)} I^{\theta}(P_1 || P_\infty) \\
\geq I^{\Theta_r(\epsilon_1)}(P_1 || P_\infty) \\
= \epsilon_1 I(P_1 || P_\infty).
$$

The proof thus is completed. 

Given $a_1$ and $\epsilon$, we define a set $E_+(a_1, \epsilon)$ as

$$
E_+(a_1, \epsilon) = \{ \epsilon_1 : \mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \right] \geq \mathbb{E}_\infty \left[ T_c(a_1) \right], \forall a > a_1 \},
$$

where $T_c(a_1)$ is the stopping time for the CuSum algorithm with $a_1$ as the threshold. Under Assumption \[1\] and the assumption that $P_\infty$ and $P_1$ are mutually locally absolutely continuous, using the standard performance analysis technique for the CuSum algorithm (e.g., P.142 of \[30\]), one can see that $\mathbb{E}_\infty \left[ T_c(a_1) \right]$ is finite for any finite $a_1$. Using the same analysis for $E(a_1, \epsilon)$ (see the proof of Proposition\[3\]), one can see that $E_+(a_1, \epsilon)$ is not empty for any $a_1$ and $\epsilon$. Furthermore, one can also obtain that

$$
E^*(a_1, \epsilon) = E(a_1, \epsilon) \cap E_+(a_1, \epsilon)
$$

(13)
is non-empty for any $a_1$ and $\epsilon$.

We are now ready to present the second main theorem.

**Theorem 2.** For any $\epsilon > 0$, when the censoring strategy $\theta^*(\epsilon_1)$ with $\epsilon_1 \in E^*(a_1, \epsilon)$ is used, the CuSum-AC algorithm with $N = 2$ levels is asymptotically ($\zeta \to \infty$) optimal for both Problems [1] and [2].

*Proof:* See Appendix D. \[\]

**Remark 8.** The interesting point of the above theorem is that to achieve the asymptotic optimality, the communication rate $r(\Theta)$ of the CuSum-AC algorithm does not necessarily equal $\epsilon$.

We then extend the above results to the CuSum-AC algorithm with $N > 2$ levels. Let

$$\tilde{T}_{c-t}(x, y) = \inf \{ k : s_k \geq y \text{ with } s_0 = x \},$$

where $s_k$ is defined in (7). Define

$$\Theta^*(\epsilon) = \left\{ \Theta : r(\Theta) \leq \epsilon, \forall a > a_1 \quad \text{and} \quad \mathbb{E}_\infty \left[ \tilde{T}_{c-t} \left( \hat{s}_{\eta(0)}, a_1 \right) \middle| \hat{s}_{\eta(0)} < a_1 \right] \geq \mathbb{E}_\infty \left[ T_c(a_1) \right], \forall a > a_1 \right\}.$$

The counterpart of $\Theta^*(\epsilon)$ for the CuSum-AC algorithm with $N = 2$ level is $E^*(a_1, \epsilon)$, which is defined in (13). Note that the $N > 2$ levels case provides more 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 $\epsilon_1$). Then following the similar arguments of $E^*(a_1, \epsilon)$, one can conclude that the set $\Theta^*(\epsilon)$ is a non-empty set.

**Theorem 3.** For any $\epsilon > 0$, when the censoring strategy $\Theta \in \Theta^*(\epsilon)$ is used, the CuSum-AC algorithm with $N > 2$ levels is asymptotically ($\zeta \to \infty$) optimal for both Problems [1] and [2].

The proof is similar to that of Theorem [2] and is omitted here.

**Remark 9.** Though the asymptotic optimality can be achieved for any censoring levels, better detection performance should be expected when $N$ increases if $\zeta$ takes moderate values.

As a corollary, we give the asymptotic performance when different censoring strategies are used: the stationary censoring strategy, the random transmission control policy and the adaptive
censoring strategy.

**Corollary 2.** For any $\epsilon$, the asymptotic performance for the CuSum algorithm with different transmission control policies are given as follows.

$$d_L(T_c, \Theta_r(\epsilon)) = d_P(T_c, \Theta_r(\epsilon)) = \frac{\ln \zeta}{\mathcal{I}(\mathcal{P}_1||\mathcal{P}_\infty)},$$

as $\zeta \to \infty$.

If the random transmission control policy $\Theta_r(\epsilon)$, which is defined in (12), is used. If the stationary censoring strategy $\theta^*(\epsilon)$ is used, as $\zeta \to \infty$,

$$\frac{\ln \zeta}{\mathcal{I}(\mathcal{P}_1||\mathcal{P}_\infty)} \leq d_L(T_c, \theta^*(\epsilon)) = d_P(T_c, \theta^*(\epsilon)) \leq \frac{\ln \zeta}{\mathcal{I}(\mathcal{P}_1||\mathcal{P}_\infty)}.$$

If adaptive censoring strategy is used, as $\zeta \to \infty$,

$$d_L(T_{c-t}, \Theta^*(\epsilon)) = d_P(T_{c-t}, \Theta^*(\epsilon)) = \frac{\ln \zeta}{\mathcal{I}(\mathcal{P}_1||\mathcal{P}_\infty)}.$$

**Proof:** This directly follows from Theorem 1, Corollary 1, Lemma 2, Theorem 3 and the well-established asymptotic performance of the CuSum algorithm (e.g., P. 160 of [30]), the details of which are omitted.

**Remark 10.** Among the three transmission control strategies: $\Theta_r(\epsilon)$, $\theta^*(\epsilon)$ and $\Theta^*(\epsilon)$, the adaptive censoring strategy $\Theta^*(\epsilon)$ is most preferred, since its asymptotic detection delay is the least, as $\zeta \to \infty$. The cost is that the feedback messages from the decision maker is required. In particular, for the CuSum-AC algorithm with $N = 2$ levels, the average number of feedback transmission under $\mathcal{P}_\infty$ is $E_\infty[N_F] = \frac{2}{1 - \mathbb{P}_\infty(\hat{s}_0(0) < a_1)}$. The average feedback transmission duty cycle is $E_\infty[N_F] \leq \frac{2}{E_\infty[\eta(0)] + E_\infty[\hat{\phi}(\hat{s}_0(0))\hat{s}_0(0) < a_1\mathbb{P}_\infty(\hat{s}_0(0) < a_1)]}$, which is related to $a_1$, $\theta^*(\epsilon)$ and $\eta$.

**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. As in the one sensor case, it assumed that $\{X_{m,1}, \ldots, X_{m,\nu-1}\}$ are i.i.d. with probability density function (pdf) $f_{\{0,m\}}$ and $\{X_{m,\nu}, \ldots\}$ are i.i.d. with pdf $f_{\{1,m\}}$. March 18, 2015 DRAFT
As in [14], [32], it is assumed that the change event affects all the sensors simultaneously at \( \nu \) and the observations are independent across the sensors, conditioned on the change point.

Like \( \gamma_k \) in [11], let \( \gamma_{m,k} \) be the indicator whether or not the sensor \( m \) sends its observation \( X_{m,k} \) to the fusion center. Let \( \theta_{m,k} \) be the censoring strategy used at the sensor \( m \) at the time instant \( k \), i.e.,

\[
\gamma_{m,k} = \theta_{m,k}(X_{m,k}).
\]

Let \( \Theta_k = \{\theta_{1,k}, \ldots, \theta_{M,k}\} \) be the censoring strategies used at all the sensor nodes at time \( k \) and \( \Theta^M \) be the censoring policy along the horizon, i.e., \( \Theta^M = \{\Theta_1, \ldots, \Theta_T\} \).

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

\[
 r(\Theta^M) = \limsup_{n \to \infty} \frac{1}{nM} E^\infty \left[ \sum_{k=1}^n \sum_{m=1}^M \gamma_{m,k} \nu \geq n \right]. \tag{14}
\]

In [23], 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. Given \( T \) and \( \Theta^M \), the ARLFA is defined in the same way as in the one sensor case, i.e.,

\[
 g(T, \Theta^M) = E^\infty [T].
\]

Let

\[
 I^m_k = \{\gamma_{m,1}, \ldots, \gamma_{m,k}, \gamma_{m,1} X_{m,1}, \ldots, \gamma_{m,k} X_{m,k}\},
\]

and \( I^M_k = \{I^1_k, \ldots, I^M_k\} \). Then the Lorden’s and Pollak’s detection delays are defined by

\[
d_L(T, \Theta^M) = \sup_{1 \leq \nu < \infty} \left\{ \text{ess sup} E_\nu [(T - \nu + 1)^+ | I^M_{\nu-1}] \right\},
\]

and

\[
d_P(T, \Theta^M) = \sup_{1 \leq \nu < \infty} E_\nu [(T - \nu + 1)^+ | T \geq \nu].
\]

Then the two problems we are interested in are as follows:
Problem 3.

\[
\begin{align*}
\text{minimize} \quad & d_L(T, \Theta^M), \\
\text{subject to} \quad & g(T, \Theta^M) \geq \zeta, \\
& r(\Theta^M) \leq \epsilon,
\end{align*}
\]

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

Problem 4.

\[
\begin{align*}
\text{minimize} \quad & d_P(T, \Theta^M), \\
\text{subject to} \quad & g(T, \Theta^M) \geq \zeta, \\
& r(\Theta^M) \leq \epsilon,
\end{align*}
\]

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

B. CuSum-AC Algorithm for Multiple Sensors Case

We only present 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

\[
T^{\text{MC}}_{c-t} = \inf \{k : s^M_k \geq a\},
\]

where the detection statistic \( s^M_k \) is updated as follows:

\[
s^M_k = \max \{0, s^M_{k-1} + \sum_{m=1}^{M} \ell^{(\theta_{\{m,k\}})}(\gamma_{\{m,k\}}, X_{\{m,k\}})\},
\]

\[
s^M_k = \begin{cases} 
   a_1, & \text{if } s^M_{k-1} < a_1 \text{ and } \tilde{s}^M_k \geq a_1, \\
   \tilde{s}^M_k, & \text{otherwise,}
\end{cases}
\]

\( s^M_0 = 0 \),

and the censoring strategies are given by

\[
\theta_{\{m,k\}} = \begin{cases} 
   \theta^*(1), & \text{if } s^M_{k-1} \geq a_1, \\
   \theta^*(\epsilon_{\{m,1\}}), & \text{otherwise,}
\end{cases}
\]

March 18, 2015
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 $X^M_k = \{X_{(1,k)}, \ldots, X_{(M,k)}\}$ and $\theta^M_k = \{\theta_{(1,k)}, \ldots, \theta_{(M,k)}\}$. Note that $X^M_k$ and $\theta^M_k$ can be regarded as the “vector version” of $X_k$ and $\theta_k$ 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 $X^M_k$ and $\theta^M_k$. Thus the following theorems are straightforward.

**Theorem 4.** The stopping time $T^M_{c-t}$ is an equalizer rule for Problem 3, i.e.,

$$d_L(T^M_{c-t}, \Theta^M) = \text{ess sup} \mathbb{E}_\nu[(T^M_{c-t} - \nu + 1)^+ | I^M_{\nu-1}], \quad \forall \nu \geq 1.$$ 

And for the stopping time $T^M_{c-t}$, the following holds:

$$d_P(T^M_{c-t}, \Theta^M) = \mathbb{E}_\nu[T^M_{c-t}].$$

**Theorem 5.** For any $\epsilon > 0$, when the censoring strategy $\Theta^M \in \Theta^M_\epsilon(\epsilon)$ is used, the CuSum-AC algorithm with $N = 2$ levels is asymptotically ($\zeta \to \infty$) optimal for both Problems 3 and 4.

Note that $E^*(a_1, \epsilon)$ is the counterpart of $\Theta^M_\epsilon(\epsilon)$ in one sensor case. To see that $\Theta^M_\epsilon(\epsilon)$ 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 $E^*(a_1, \epsilon)$.

**VI. Numerical Examples**

For simulations, we consider the problem of mean shift detection in Gaussian noise. Except in Example 3, 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 all the examples, the CuSum-AC algorithm with $N = 2$ levels is used and in each example, the sensors use an identical censoring strategy.

**Example 1.** We illustrate the results that the CuSum-AC algorithm with $N = 2$ levels is an equalizer rule for Problem 3 and the associated detection delay for Problem 4 is attained when the change event happens at $\nu = 1$. Let $d^L_L(T, \Theta^M) = \text{ess sup} \mathbb{E}_\nu[(T - \nu + 1)^+ | I^M_{\nu-1}]$ and $d^P_P(T, \Theta^M) = \mathbb{E}_\nu[(T - \nu + 1)^+ | T \geq \nu]$. For simulations, the parameters for the CuSum-AC algorithm with 2 levels are as follows: $a = 7, a_1 = 3$ and $\epsilon_{m,1} = 0.5$. Both $d^L_L(T, \Theta^M)$ and $d^P_P(T, \Theta^M)$...
Fig. 6: Quantities $d_L^\nu(T, \Theta^M)$ and $d_P^\nu(T, \Theta^M)$ as functions of the time instant the change event happens.

and $d_P^\nu(T, \Theta^M)$ are computed with the change time $\nu$ varying from 1 to 50. As illustrated in Fig. 6, $d_L^\nu(T, \Theta^M)$ remains the same when $\nu$ varies (the fluctuation of the values is caused by the inherent error of Monte Carlo experiments), whereas $d_P^1(T, \Theta^M)$ is the largest one among \{ $d_P^1(T, \Theta^M), \ldots, d_P^{50}(T, \Theta^M)$ \}. This is precisely the idea shown in Theorem 4.

**Example 2.** 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 (which is the optimal one when there is no communication rate constraint). With different ARLFA’s (sufficiently large), the detection delays (i.e., $E_1[T]$) 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., $a, a_1, \epsilon_1$). 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 $\epsilon_1$, 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 $\epsilon_1$, the communication rate remains the same when $a$ varies if $a$ is sufficiently large. The simulation results are given in Fig. 7. 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 trade-off for the CuSum-AC algorithm, which will be shown further in the next example.

**Example 3.** We compare the CuSum-AC algorithm with the CuSum-SC algorithm, the asymptotically optimal one can only be found for the one sensor case. One sensor thus is used and it is assumed that \( f_0 \sim \mathcal{N}(0, 1) \) and \( f_1 \sim \mathcal{N}(1, 1) \). We plot Fig. 8 in the following way. The ARLFA is fixed to be 70000, i.e., the parameters for the algorithm ( \( a \) for the CuSum algorithm with stationary censoring strategy and random transmission control policy, and \( a, a_1 \) and \( \epsilon_1 \) for the CuSum-AC algorithm) should be chosen such that the associated ARLFA is around 70000. The stationary censoring strategy can be determined by the communication rate (see our previous work [29]), and given the stationary censoring strategy and the required ARLFA, the threshold \( a \) for the CuSum algorithm is obtained by the bi-section technique. The parameters for the CuSum-AC algorithm are determined using the brute-force search technique. The admissible combinations of the parameters \( a, a_1 \) and \( \epsilon_1 \) are multiple, among which the one that has the smallest detection delay is used. As depicted in Fig. 8, for any communication rate among \([0.1, 1]\), the CuSum-AC algorithm has smaller detection delay than that of the CuSum-SC algorithm.
Fig. 8: Detection delay versus the communication rate for the CuSum algorithm with different transmission control policy.

In particular, the detection delay of the CuSum-AC algorithm is only 2.4 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). Compared with the random transmission control policy, the censoring strategy significantly reduces the detection delay, in particular when the allowed communication rate is small.

Example 4. In this example, the hard thresholding scheme $N_{\text{hard}}$ in [15] is used to compare with the CuSum-AC algorithm. We remark that the comparison might be unfair. On one hand, the $N_{\text{hard}}$ algorithm works in scenarios where the remote sensors have sufficient computation capability, while the CuSum-AC algorithm is for sensors that have limited computation capability. On the other hand, the CuSum-AC algorithm needs feedback information from the fusion center, while this is not the case for the $N_{\text{hard}}$ algorithm. Since the sensors are identical, we let the sensors use the same local threshold for the $N_{\text{hard}}$. For simulations, the ARLFA is around 10000. The global threshold at the fusion center is around 11. The communication rate constraint is satisfied by adjusting both the local threshold and the global threshold (they are coupled). The result is illustrated in Fig. 9, from which one can see that the detection delay of the CuSum-AC algorithm is larger than that of the $N_{\text{hard}}$ algorithm when $\epsilon = 0.1$ but much smaller when $\epsilon \geq 0.2$. 

March 18, 2015
VII. Conclusion and Future Work

In this paper, we have studied the problem of non-Bayesian quickest change detection with communication rate constraints. The constraint is posed by limited bandwidth of the communication channels between the remote sensor networks and the fusion center. Two minimax formulations for this problem are proposed, which are extensions of the classical Lorden’s and Pollak’s formulations. 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 is proved to be an equalizer rule for the Lorden’s criterion, and be globally asymptotically optimal for any communication rate constraint, as the ARLFA goes to infinity, for both problems. The numerical simulations show that the CuSum-AC algorithm has a better detection performance versus communication rate trade-off than the CuSum algorithm coupled with the optimal stationary censoring strategy and the CuSum algorithm with random transmission control policy. It is also shown, via numerical simulations, the detection delay of the CuSum-AC algorithm is less than that of the \( N_{\text{hard}} \) algorithm in the existing literature when the communication resources are not that severe.

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 multi-channel setting (the change event only affects a subset of the sensors).
APPENDIX

A. Proof of Lemma 2

We first prove that for any $0 \leq z < a_1$,

$$E_1 [T_{c-l}(z)] \leq E_1 [T_{c-l}]$$ \hspace{1cm} (17)

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

$$E_1 [T_{c-l}(z)] = E_1 [\phi(z)] + E_1 [T_{c-l}(a_1)].$$ \hspace{1cm} (18)

The quantity $E_1 [T_{c-l}(a_1)]$ is a common term. Thus to obtain (17), we only need to prove that for any $0 \leq z < a_1$

$$E_1 [\phi(z)] \leq E_1 [\phi(0)].$$ \hspace{1cm} (19)

Given the censoring strategy $\theta^*(\epsilon_1)$, for each possible sample path $\{X_k\}$, the quantity $\hat{s}_k(z)$ (which is defined in (9)) is always no less than $\hat{s}_k(0)$. The equation (19) thus follows.

We then prove that for any $a_1 \leq z < a$,

$$E_1 [T_{c-l}(z)] \leq E_1 [T_{c-l}]$$ \hspace{1cm} (20)

In the remainder of this proof, without explicit statements, the variable $z$ is assumed to be in the range $[a_1, a)$. 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., $s_k(z) \geq a_1$ along the whole horizon; we denote this event by $Z_1^\uparrow$. The other one is that the detection statistic $s_k(z)$ once crosses $a_1$ from up before the algorithm stops, which is denoted by $Z_2^\uparrow$. Let

$$p(z) = P_1 \{ Z_2^\uparrow \}.$$ \hspace{1cm}

We then have

$$E_1 [T_{c-l}(z)] = p(z) E_1 [T_{c-l}(z) | Z_2^\uparrow] + (1 - p(z)) E_1 [T_{c-l}(z) | Z_1^\uparrow]$$

$$= p(z) E_1 [T_{c-l}(z) | Z_2^\uparrow]
+ (1 - p(z)) (E_1 [\eta(z) | Z_2^\uparrow] + E_1 [T_{c-l}(x); \hat{s}_{\eta(z)} = x | Z_2^\uparrow])$$

$$:= p(z) t_1^\uparrow + (1 - p(z)) \left( t_1^\uparrow + t_2^\uparrow \right).$$ \hspace{1cm} (21)
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 $z$.

To obtain (20) based on (21), we first define a stopping time $\tilde{T}_c$ as follows. This is a stopping time for an SPRT that works in the same manner with the CuSum algorithm but starts at $a_1$ and bounded below by $a_1$:

$$\tilde{T}_c = \inf\{k : \tilde{c}_k \geq a\},$$

where $\tilde{c}_k$ involves by

$$\tilde{c}_k = \max(\tilde{c}_{k-1} + \ell(X_k), a_1),$$

$$\tilde{c}_0 = a_1.$$

For the quantities $p(z), t^\uparrow$ and $t_1^\downarrow$, from Lemma 5 of [20], we then have an inequality as follows. For any $t^* \geq \mathbb{E}_1[\tilde{T}_c]$,

$$p(z)t^\uparrow + (1-p(z))(t_1^\downarrow + t^*) \leq t^*.$$

(22)

Because of the reset action for the CuSum-AC algorithm when crossing $a_1$ from below and the fact that $\{X_k\}'s$ are i.i.d under the measure $\mathbb{P}_1$, the following can be obtained:

$$\mathbb{E}_1 [T_{c-l}(a_1)] \geq \mathbb{E}_1[\tilde{T}_c].$$

(23)

Combining (17), (18) and (23), one can obtain:

$$\mathbb{E}_1 [T_{c-l}] \geq \mathbb{E}_1[\tilde{T}_c].$$

(24)

We then study the quantity $t_2^\downarrow$.

$$t_2^\downarrow = \mathbb{E}_1 [T_{c-l}(x); \hat{s}_{\eta(z)} = x| Z_x^\downarrow]$$

$$= \int_{0 \leq x < a_1} \mathbb{E}_1 [T_{c-l}(x)| Z_x^\downarrow; \hat{s}_{\eta(z)} = x] \, d\mathbb{P}_1\{\hat{s}_{\eta(z)} \leq x| Z_x^\downarrow\}$$

$$\leq \esssup_{0 \leq x < a_1} \mathbb{E}_1 [T_{c-l}(x)]$$

$$\leq \mathbb{E}_1 [T_{c-l}].$$

(25)
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. Note that \(\int_{0 \leq x < a_1} d\mathbb{P}_1 \{ \hat{s}_{\eta(z)} \leq x \} = 1\). The inequality \((ie2)\) follows directly from \((17)\).

We are ready to prove \((20)\).

\[
\mathbb{E}_1 [T_{c-t}(z)] = p(z) t^t + (1 - p(z)) \left( t^t_1 + t^t_2 \right)
\]

\[
\overset{(ie1)}{\leq} p(z) t^t + (1 - p(z)) \left( t^t_1 + \mathbb{E}_1 [T_{c-t}] \right)
\]

\[
\overset{(ie2)}{\leq} \mathbb{E}_1 [T_{c-t}],
\]

where the inequality \((ie1)\) follows from \((25)\) and \((ie2)\) follows from \((22)\) and \((24)\).

The equation \((11)\) follows directly from \((17)\) and \((20)\). The proof thus is completed.

B. Proof of 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.,

\[
\mathbb{E}_{\nu}[ (T_{c-t} - \nu + 1)^+ | T_{\nu-1} ] = \mathbb{E}_{\nu}[ (T_{c-t} - \nu + 1)^+ | s_{\nu-1} ] = \mathbb{E}_{\nu}[ (T_{c-t} | s_{\nu-1} )].
\]

Note that under Assumption \(1\) for any censoring strategy \(\theta^*(\epsilon_1)\), from the positive definiteness of K-L divergence \([33]\), one can have

\[
\mathbb{E}_\infty \left[ \ell^{\theta^*(\epsilon_1)}(\gamma_k, X_k) \right] < 0.
\]

We then have

\[
\mathbb{P}_\infty \{ \ell^{\theta^*(\epsilon_1)}(\gamma_k, X_k) < 0 \} := p_{x_-} > 0.
\]

The following thus holds:

\[
\mathbb{P}_\infty \{ s_{\nu-1} = 0 \} > p_{x_-}^{\nu-1} > 0.
\]

(26)
We then can obtain:

\[ \text{ess sup } E_{\nu}[(T_{c-t} - \nu + 1)^+|I_{\nu-1}] \]
\[ = \text{ess sup } E_{\nu}[T_{c-t}(s_{\nu-1})] \]
\[ = E_{\nu}[T_{c-t}(s_{\nu-1} = 0)], \]

where the last equality follows from Lemma 1 and (26). The proof thus is completed.

C. Proof of Proposition 7

Recall that in Section IV-A the CuSum-AC algorithm is interpreted as a sequence of SPRTs. The same technique is used in this proof. We first give the upper bound of \( r(\Theta) \) when \( a, a_1 \) and \( \theta^*(\epsilon_1) \) are known. We then show that by choosing certain \( \epsilon_1 \) (equivalently with \( \theta^*(\epsilon_1) \)), which is independent of \( a \), the upper bound can be any admissible value.

Define a random sequence \( \{T_i\} \) as

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

where \( W_i = W^1_i + W^2_i \) with \( W^1_i \) i.i.d. distributed with mean equals \( E_\infty [\eta(0) | \hat{s}_{\eta(0)} < a_1] \) and \( W^2_i \) i.i.d. distributed with mean equals \( E_\infty [\phi(\hat{s}_{\eta(0)}) | \hat{s}_{\eta(0)} < a_1] \). Note that the distribution of \( W_i \) is different from that of \( \Phi \) defined in (10). The stopping time \( \Phi \) 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

\[ R_0 = 0, \]
\[ R_i = T_{i-1} + W^1_i + J_i, \quad \forall i \geq 1, \]

where \( J_i \sim B(W^2_i, \epsilon_1) \) is a binomial distributed random variable.
Given $\epsilon_1, a_1$ and $a$,

$$r(\Theta) \leq \limsup_{n \to \infty} \frac{1}{n - T_0} \mathbb{E}_\infty \left[ \sum_{k=T_0+1}^{n} \gamma_k \mid \nu > n \right]$$

$$= \lim_{i \to \infty} \frac{R_i}{T_i - T_0}$$

$$= \epsilon_1 \frac{\mathbb{E}_\infty [W_1^1 + J_i]}{\mathbb{E}_\infty [W_i]}$$

$$= \epsilon_1 \frac{\mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \mid \hat{s}_{\eta(0)} < a_1 \right] + \mathbb{E}_\infty \left[ \eta(0) \mid \hat{s}_{\eta(0)} < a_1 \right]}{\mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \mid \hat{s}_{\eta(0)} < a_1 \right] + \mathbb{E}_\infty \left[ \eta(0) \mid \hat{s}_{\eta(0)} < a_1 \right]}$$

$$:= \bar{r}(\Theta),$$

where the inequality (ie1) holds because before the time instant $T_0$, all the observations are censored using the censoring strategy $\theta^*(\epsilon_1)$ (from the definition of $\phi(0)$), the communication rate of which is $\epsilon_1$ (the lower bound of $r(\Theta)$); the equality (e1) follows from the renewal reward theory [34].

Using the sample path analysis, one can see that, given $a_1$, $\mathbb{E}_\infty \left[ \eta(0) \mid \hat{s}_{\eta(0)} < a_1 \right]$ is monotonically increasing (non-decreasing) with $a$. Let

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

$$= \mathbb{E}_\infty \left[ \eta(0) \mid a = +\infty \right],$$

where the second equality follows because when $a = +\infty$, the event $\{\eta(0) \geq 1\}$ is equivalent with $\{s_{\eta(0)} < a_1\}$. As $0 < I(\mathbb{P}_\infty || \mathbb{P}_1) < \infty$ (from Assumption [1]), $\mathbb{E}_\infty \left[ \ell(X_k) \right] < 0$. From Corollary 2.4 of [35], one can see that $T_{a_1}^\infty$ is finite.

Since $\bar{r}(\Theta)$ is monotonically increasing with $\mathbb{E}_\infty \left[ \eta(0) \mid \hat{s}_{\eta(0)} < a_1 \right]$, one can obtain that for any $a$

$$\bar{r}(\Theta) \leq \epsilon_1 \frac{\mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \mid \hat{s}_{\eta(0)} < a_1 \right]}{\mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \mid \hat{s}_{\eta(0)} < a_1 \right]} + \frac{T_{a_1}^\infty}{T_{a_1}^\infty}$$

$$= \epsilon_1 + \frac{\mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \mid \hat{s}_{\eta(0)} < a_1 \right]}{1 + \mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) \mid \hat{s}_{\eta(0)} < a_1 \right]}$$

$$:= \bar{r}^+(\Theta).$$
For any censoring strategy $\theta^*(\epsilon_1)$, $\mathbb{E}_\infty [\ell^{\theta^*(\epsilon_1)}(\gamma_k, X_k)] \leq 0$. Furthermore, as $\epsilon_1 \to 0$, $\mathbb{E}_\infty [\ell^{\theta^*(\epsilon_1)}(\gamma_k, X_k)] \to 0$. 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. From Corollary 2.6 of [35], one can see that as $\mathbb{E}_\infty [\ell^{\theta^*(\epsilon_1)}(\gamma_k, X_k)] \to 0$, $\mathbb{E}_\infty [\phi(z)] \to \infty$ 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 $\theta^*(\epsilon_1)$ (i.e., $\epsilon_1$). Given any $\epsilon$, one thus can find a non-empty set $E(a_1, \epsilon)$ that is independent of $a_1$ (i.e., independent of the distribution of $\{\hat{s}_{\eta(0)}|\hat{s}_{\eta(0)} < a_1\}$) such that for any censoring strategy $\theta^*(\epsilon_1)$ with $\epsilon_1 \in E(a_1, \epsilon)$,

$$\bar{r}^+(\Theta) \leq \epsilon.$$

The proof thus is completed.

D. Proof of Theorem 2

Note that the CuSum algorithm is strictly optimal for Problem 1 and asymptotically optimal for Problem 2 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. To this end, we define a stopping time as follows:

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

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

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

1 This approach is not necessarily monotonic. In some cases, $\mathbb{E}_\infty [\ell^{\theta^*(\epsilon_1)}(\gamma_k, X_k)]$ can be zero for non-zero $\epsilon_1$’s.
The difference between $\hat{c}_k$ and $c_k$ in (5) for the CuSum algorithm is that $\hat{c}_k$ is reset to be $a_1$ whenever it crosses $a_1$ from below (up). Using the sample path technique, one can see that for any $a$,

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

(27)

Let $\hat{C}_c(k)$ be the counter indicating the number of $\hat{c}_k$ crossing $a_1$ from below before time instant $k$, i.e,

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

with $\hat{C}_c(0) = 0$. Wald’s identity [36] yields

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

(28)

Let

$$
T_{c_{-\ell}}(a) = \inf \{ k : s_k \geq a \},
$$

where $s_k$ is evolved in the same manner with $s_k$ the stopping time $T_{c_{-\ell}}$ when $N = 2$ except for the starting point. The detection statistic $s_k$ starts at $s_0 = 0$, while $\hat{s}_k$ starts with a random variable $\hat{s}_k = x$ and the distribution of $x$ is the same with the conditional variable $\{ \eta(0) | \hat{s}_{\eta(0)} < a_1 \}$.

Using the sample path technique, one can see that

$$
\mathbb{E}_\infty [T_{c_{-\ell}}] \geq \mathbb{E}_\infty [T_{c_{-\ell}}(a)].
$$

(29)

Let $C_{c_{-\ell}}(k)$ be the counter indicating the number of $\hat{s}_k$ crossing $a_1$ from below before time instant $k$, which is defined in the same manner with $\hat{C}_c(k)$. Also by Wald’s identity, one can have

$$
\mathbb{E}_\infty \left[ T_{c_{-\ell}}(a) \right] = \left( \mathbb{E}_\infty [\eta(0)] + \mathbb{E}_\infty \left[ \phi \left( \hat{s}_{\eta(0)} \right) | \hat{s}_{\eta(0)} < a_1 \right] \mathbb{P}_\infty \{ \hat{s}_{\eta(0)} < a_1 \} \right) \mathbb{E}_\infty \left[ C_{c_{-\ell}} \left( T_{c_{-\ell}}(a) \right) \right].
$$

(30)
Both $\hat{c}_k$ and $s_k$ are reset to be $a_1$ when they cross $a_1$ from below, we thus obtain the following:

$$E_\infty \left[ \hat{C}_c \left( \hat{T}_c(a) \right) \right] = E_\infty \left[ C_{c-t} \left( T_{c-t}(a) \right) \right] = \frac{1}{1 - \mathbb{P}_\infty \{ \hat{s}_n(0) < a_1 \}}, \quad \text{(31)}$$

where the last equality follows because both $\hat{C}_c(k)$ and $C_{c-t}(k)$ are geometrically distributed.

Combining (13), (27), (28), (29), (30) and (31), one can see that for any $\epsilon$, when the censoring strategy $\theta^*(\epsilon_1)$ with $\epsilon_1 \in E^*(a_1, \epsilon)$ is used

$$E_\infty [T_{c-t}] \geq E_\infty [T_c(a)]. \quad \text{(32)}$$

**Step 2.** We show that for any $\epsilon$, as $a \to \infty$,

$$E_1 [T_{c-t}] = E_1 [T_c(a)] (1 + o(1)).$$

Let

$$T_{a_1+} = \sum_{k=1}^{T_{c-t}} 1_{\{s_k \geq a_1\}},$$

$$T_{a_1-} = T_{c-t} - T_{a_1+}.$$

Recall that $s_k$ is the detection statistic for the stopping time $T_{c-t}$ when $N = 2$. The quantity $T_{a_1+}$ ($T_{a_1-}$) can be viewed as the duration that $s_k$ stays above (below) $a_1$. Following similar arguments in **Step 1**, one can obtain

$$T_{a_1+} = E_1 [\eta(0)] \frac{1}{1 - \mathbb{P}_1 \{ \hat{s}_n(0) < a_1 \}},$$

$$T_{a_1-} \leq E_1 [\phi(0)] \frac{1}{1 - \mathbb{P}_1 \{ \hat{s}_n(0) < a_1 \}}.$$

Since $I(\mathbb{P}_1 || \mathbb{P}_\infty) > 0$, from Corollary 2.4 of [35], one can obtain that

$$E_1 [\eta(0)] \to \infty, \quad \text{as } a \to \infty.$$

From Lemma 2 for any $\epsilon_1 > 0$, $I^{\theta^*(\epsilon_1)}(\mathbb{P}_1 || \mathbb{P}_\infty) > 0$. From the established performance analysis technique for the CuSum algorithm (e.g., P. 142 of [30]), one can see that

$$E_1 [\phi(0)] < \infty.$$
Note that $E_1[\phi(0)]$ is only related to $a_1$ and $\theta^*(\epsilon_1)$. From the definition of $E^*(a_1, \epsilon)$, $\theta^*(\epsilon_1)$ is independent of $a$. The following thus can be obtained:

$$\frac{T_{a_1} - T_{a_1^+}}{E_1[\phi(0)]} \leq \frac{E_1[\phi(0)]}{E_1[\eta(0)]} \to 0, \quad \text{as } a \to \infty.$$  

Then $a \to \infty$,

$$E_1[T_{c-t}] = T_{a_1^+}((1 + o(1))). \quad (33)$$

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

$$T_{a_1^+} \leq E_1[T_c(a - a_1)] = \frac{\ln(a - a_1)}{I(P_1||P_\infty)}, \quad \text{as } a \to \infty. \quad (34)$$

Note that as $a \to \infty$

$$E_1[T_c(a)] = \frac{\ln(a)}{I(P_1||P_\infty)}. \quad (35)$$

Combining (33), (34) and (35), one can obtain that for any $\epsilon$, as $a \to \infty$,

$$E_1[T_{c-t}] \leq E_1[T_c(a)] (1 + o(1)) \overset{(e1)}{=} E_1[T_c(a)] (1 + o(1)), \quad (36)$$

where (e1) holds because given (32), $E_1[T_c(a)]$ is the lower bound of $E_1[T_{c-t}]$ as $a \to \infty$.

**Step 3.** From Theorem 1 and Corollary 1 one can see that

$$d_L(T_{c-t}, \Theta) = d_P(T_{c-t}, \Theta) = E_1[T_{c-t}].$$

Note that $T_c(a)$ is asymptotically optimal for both Problems 1 and 2 as $\zeta \to \infty$ (i.e., $a \to \infty$), when $\epsilon = 1$. Combining (32) and (36), one can obtain the statement in Theorem 2. The proof thus is completed.

**REFERENCES**

[1] J. A. Rice, K. Mechitov, S.-H. Sim, T. Nagayama, S. Jang, R. Kim, B. F. Spencer Jr, G. Agha, and Y. Fujino, “Flexible smart sensor framework for autonomous structural health monitoring,” *Smart structures and Systems*, vol. 6, no. 5-6, pp. 423–438, 2010.

[2] 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.
[3] T. L. Lai, “Sequential changepoint detection in quality control and dynamical systems,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 613–658, 1995.

[4] 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 Transactions on Signal Processing*, vol. 54, no. 9, pp. 3372–3382, 2006.

[5] L. Lai, Y. Fan, and H. V. Poor, “Quickest detection in cognitive radio: A sequential change detection framework,” in *Global Telecommunications Conference (GLOBECOM)*, 2008, pp. 1–5.

[6] K. W. Choi, W. S. Jeon, and D. G. Jeong, “Sequential detection of cyclostationary signal for cognitive radio systems,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4480–4485, 2009.

[7] A. Shiryaev, “On optimum methods in quickest detection problems,” *Theory of Probability and Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[8] A. N. Shiryaev, *Optimal stopping rules*. Springer, 2007.

[9] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[10] M. Pollak, “Optimal detection of a change in distribution,” *The Annals of Statistics*, pp. 206–227, 1985.

[11] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions;” *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[12] Y. Ritov, “Decision theoretic optimality of the cusum procedure,” *The Annals of Statistics*, pp. 1464–1469, 1990.

[13] 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.

[14] V. Veeravalli, “Decentralized quickest change detection,” *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[15] Y. Mei, “Quickest detection in censoring sensor networks,” in *2011 IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2011, pp. 2148–2152.

[16] 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.

[17] K. Premkumar and A. Kumar, “Optimal sleep-wake scheduling for quickest intrusion detection using wireless sensor networks,” in *Proceedings of the International Conference on Computer Communications IEEE (INFOCOM)*, 2008, pp. 1400–1408.

[18] 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.

[19] J. Geng, E. Bhayraktar, and L. Lai, “Bayesian quickest change point detection with sampling right constraints,” *arXiv preprint arXiv:1309.5396*, 2013.

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

[21] L. Zacharias and R. Sundaresan, “Decentralized sequential change detection using physical layer fusion,” *Wireless Communications, IEEE Transactions on*, vol. 7, no. 12, pp. 4999–5008, 2008.

[22] T. Banerjee, V. Sharma, V. Kavitha, and A. Jayaprakasam, “Generalized analysis of a distributed energy efficient algorithm for change detection,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 91–101, 2011.
[23] T. Banerjee and V. V. Veeravalli, “Data-efficient quickest outlying sequence detection in sensor networks,” arXiv preprint arXiv:1411.0183, 2014.

[24] W. Dargie and C. Poellabauer, Fundamentals of wireless sensor networks: theory and practice. John Wiley & Sons, 2010.

[25] B. E. B. K. P. L. Doherty, B. A. Warneke, “Energy and performance considerations for smart dust,” International Journal on Parallel and Distributed Systems and Networks, vol. 4, no. 3, pp. 121–133, 2001.

[26] C. Rago, P. Willett, and Y. Bar-Shalom, “Censoring sensors: A low-communication-rate scheme for distributed detection,” IEEE Transactions on Aerospace and Electronic Systems, vol. 32, no. 2, pp. 554–568, 1996.

[27] T. Banerjee and V. V. Veeravalli, “Data-efficient quickest change detection in distributed and multi-channel systems,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013. IEEE, 2013, pp. 3952–3956.

[28] A. G. Tartakovsky, M. Pollak, and A. S. Polunchenko, “Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures,” Theory of Probability & Its Applications, vol. 56, no. 3, pp. 457–484, 2012.

[29] X. Ren, J. Chen, K. H. Johansson, and L. Shi, “Quickest change detection with a censoring sensor in the minimax setting,” arXiv preprint arXiv:1411.3107, 2014.

[30] H. V. Poor and O. Hadjiliadis, Quickest detection. Cambridge University Press Cambridge, 2009, vol. 40.

[31] S. Kullback, Information theory and statistics. Courier Dover Publications, 1978.

[32] A. G. Tartakovsky and V. V. Veeravalli, “Asymptotically optimal quickest change detection in distributed sensor systems,” Sequential Analysis, vol. 27, no. 4, pp. 441–475, 2008.

[33] T. M. Cover and J. A. Thomas, Elements of information theory. John Wiley & Sons, 2006.

[34] S. Asmussen, Applied probability and queues. Springer, 2003, vol. 51.

[35] M. Woodroofe, Nonlinear renewal theory in sequential analysis. Siam, 1982, vol. 39.

[36] D. Siegmund, Sequential analysis: tests and confidence intervals. Springer, 1985.