ISIT'2013 1569720097

Decentralized Data-Efficient Quickest Change Detection

Taposh Banerjee, Venugopal V. Veeravalli

Dept. of Electrical & Computer Eng.
University of Illinois, Urbana-Champaign
Urbana, USA
Email: banerje5,vvv@illinois.edu

Alexander Tartakovsky
Department of Mathematics
and
Center for Applied Mathematical Sciences
University of Southern California
Los Angeles, USA
Email: tartakov@usc.edu

***Abstract*—The problem of decentralized quickest change detection is studied with an additional constraint on the cost of observations used at each sensor. Minimax problem formulations are proposed for the problem. A distributed algorithm called the DE-All algorithm is proposed in which on-off observation control is employed locally at each sensor. It is shown that the proposed algorithm is asymptotically optimal up to first order for the proposed formulations.**

I. Introduction

In many engineering applications, a sensor network is deployed for the purposes of statistical inference. One important application is the detection of an abrupt change in the statistical properties of a phenomenon under observation. Examples include detecting a sudden stress in infrastructure, detecting the arrival of a species to a habitat in habitat monitoring applications, surveillance for security related applications, detection of a primary radio in cognitive radio networks, environment monitoring, quality control, *etc.*; see [1]. In this paper we study the abrupt change detection problem in the framework of decentralized quickest change detection introduced in [2], and further studied in [3] and [4].

In the problem of decentralized quickest change detection, there are multiple sensors, and a sequence of random variables is observed at each sensor. At each time, a processed version of the observations is transmitted from each sensor to a common node, called the *fusion center*. At an unknown point in time, the *change point*, the distribution of the observations at all the sensors changes. The objective is to find an efficient way to process the observations at the sensors and to find a stopping time based on the information received by the fusion center, so as to detect the change in distribution as quickly as possible, subject to a constraint on the false alarm rate.

A sensor network is a resource constrained system, and there is a cost associated with acquiring data at the sensors, and there is also a cost associated with the communication between the sensors and the fusion center. In [2], [3] and [4], the cost of communication is controlled using the techniques of quantization or censoring. However, the cost of observations used at the sensors is not taken into account.

In this paper, we propose a distributed algorithm called the *DE-All algorithm*, in which on-off observation control ('sleep-wake' scheduling) is employed locally at each sensor to control the cost of observations used at the sensors. Since, reducing the cost of communication is also important, the algorithm employs censoring to control the cost of communication. At the fusion center the information transmitted from the sensors is fused to detect the change. We show that the DE-All algorithm is asymptotically first-order minimax optimal for minimax criteria introduced in this paper. In the minimax formulations, the objective is to find an observation control policy and a transmission policy at the sensors, and a stopping time on the received information at the fusion center, so as to minimize some version of the worst case average detection delay. This delay has to be minimized subject to constraints on the false alarm rate and the fraction of time observations are taken before the change point at each sensor.

The observation control in the DE-All algorithm is based on the DE-CuSum algorithm proposed in [5] to detect a change in a single sequence of random variables. In [5], the minimax formulations of [6] and [7] have been extended by putting an additional constraint on the cost of observations used before the change point. Since the DE-CuSum algorithm is asymptotically first-order minimax in a single sensor scenario, as shown in [5], it is a natural candidate for a distributed multi-sensor scenario considered in this paper.

The paper is organized as follows. In Section II we discuss the problem formulations. In Section III we propose the DE-All algorithm and prove its asymptotic optimality. In Section V we compare the performance of the DE-All algorithm with two existing schemes.

II. Problem formulation

In the sensor network model we consider, there are L sensors and a central decision maker called the fusion center. The sensors are indexed by the index $\ell \in \{1, \cdots, L\}$. In the following we say sensor ℓ to refer to the sensor indexed by ℓ. At sensor ℓ the sequence $\{X_{n,\ell}\}_{n\geq 1}$ is observed, where n is the time index. At some unknown time γ, $\gamma = 0, 1, 2, \ldots$, the distribution of $\{X_{n,\ell}\}$ changes from $f_{0,\ell}$ to say $f_{1,\ell}$. We assume that $\{X_{n,\ell}\}$ are independent across the indices n and ℓ conditioned on γ.

In the following, we use $\mathbb{E}_k$, $\mathbb{P}_k$ to denote the expectation and the probability measure when the change occurs at time

$\gamma = k$, $k < \infty$ and $\mathbb{E}_\infty$, $\mathbb{P}_\infty$ correspond to the case where there is never a change. We say $p(\alpha) \sim q(\alpha)$ or $p(\alpha) \leq q(\alpha)(1 + o(1))$, as $\alpha \to 0$, to denote $p(\alpha)/q(\alpha) \to 1$ and $\lim_\alpha p(\alpha)/q(\alpha) \leq 1$, respectively, as $\alpha \to 0$. We use $D(f \,||\, g)$ to represent the Kullback–Leibler (K-L) divergence between the probability density functions (p.d.fs) f and g. We assume that the moments of up to third order of all the log likelihood ratios appearing in this paper are finite and positive.

We consider the following class of control policies. At sensor ℓ, at each time $n, n \geq 0$, a decision is made as to whether to *take* or *skip* the observation at time $n + 1$. Let $M_{n,\ell}$ be the indicator random variable such that

$$M_{n,\ell} = \begin{cases} 1 & \text{if } X_{n,\ell} \text{ is used for decision making at sensor } \ell \\ 0 & \text{otherwise .} \end{cases}$$

Let $\phi_{n,\ell}$ be the observation control law at sensor ℓ, i.e.,

$$M_{n+1,\ell} = \phi_{n,\ell}(I_{n,\ell}),$$

where

$$I_{n,\ell} = \left(M_{1,\ell}, \dots, M_{n,\ell}, X_{1,\ell}^{(M_{1,\ell})}, \dots, X_{n,\ell}^{(M_{n,\ell})}\right).$$

Here, $X_{n,\ell}^{(M_{n,\ell})} = X_{1,\ell}$ if $M_{1,\ell} = 1$, otherwise $X_{1,\ell}$ is absent in the information vector $I_{n,\ell}$. Thus, the decision to take or skip a sample at sensor ℓ is based on its past information. Let

$$Y_{n,\ell} = g_{n,\ell}(I_{n,\ell})$$

be the information transmitted from sensor ℓ to the fusion center. If no information is transmitted to the fusion center, then $Y_{n,\ell}$ = NULL, which is treated as zero at the fusion center. Here, $g_{n,\ell}$ is the transmission control law at sensor ℓ. Let

$$\mathbf{Y}_n = (Y_{n,1}, \cdots, Y_{n,L})$$

be the information received at the fusion center at time n, and let τ be a stopping time on the sequence $\{\mathbf{Y}_n\}$.

Let

$$\boldsymbol{\phi_n} = (\phi_{n,1}, \cdots, \phi_{n,L}),$$

denote the observation control law at time n, and let

$$\mathbf{g}_n = (g_{n,1}, \cdots, g_{n,L}),$$

denote the transmission control law at time n. For decentralized data-efficient quickest change detection we consider the policy of type Π defined as

$$\Pi = \{\tau, (\boldsymbol{\phi}_0, \cdots, \boldsymbol{\phi}_{\tau-1}), (\mathbf{g}_1, \cdots, \mathbf{g}_\tau)\}.$$

To capture the cost of observations used at each sensor before change, we use the Pre-Change Duty Cycle (PDC) metric introduced in [5]. The PDC_ℓ, the PDC for sensor ℓ, is defined as

$$\mathsf{PDC}_\ell(\Pi) = \limsup_{\gamma \to \infty} \frac{1}{\gamma} \mathbb{E}_\gamma \left[\sum_{k=1}^{\gamma-1} M_{k,\ell} \middle| \tau \geq \gamma \right]. \tag{1}$$

Thus, PDC_ℓ is the fraction of time observations are taken before change at sensor ℓ. If all the observations are used at sensor ℓ, then $\mathsf{PDC}_\ell = 1$. If every second sample is skipped at sensor ℓ, then $\mathsf{PDC}_\ell = 0.5$.

We now propose data-efficient extensions of the formulations in [6] and [7] for decentralized quickest change detection. For data-efficient extension of Lorden's formulation [6], we consider the Worst case Average Detection Delay (WADD) metric

$$\mathsf{WADD}(\Pi) = \sup_\gamma \text{ess sup}\, \mathbb{E}_\gamma \left[(\tau - \gamma)^+ | \mathbf{Y_1}, \cdots, \mathbf{Y_{\gamma-1}} \right], \tag{2}$$

where $(x)^+ = \max\{x, 0\}$, and the False Alarm Rate (FAR) metric

$$\mathsf{FAR}(\Psi) = 1/\mathbb{E}_\infty[\tau] \tag{3}$$

expressed via the average run length (ARL) to false alarm $\mathbb{E}_\infty[\tau]$. The data-efficient extension of the formulation in [6] is

Problem 1:

$$\begin{aligned} \underset{\Pi}{\text{minimize}} \quad & \mathsf{WADD}(\Pi), \\ \text{subject to} \quad & \mathsf{FAR}(\Pi) \leq \alpha, \\ & \mathsf{PDC}_\ell(\Pi) \leq \beta_\ell, \quad \text{for} \quad \ell = 1, \cdots, L, \end{aligned} \tag{4}$$

where $0 \leq \alpha, \beta_\ell \leq 1$, $\ell = 1, \cdots, L$, are given constraints.

We also consider the minimax extension of Pollak's formulation [7], where instead of WADD, the following maximal Conditional Average Detection Delay (CADD) metric is used:

$$\mathsf{CADD}(\Psi) = \sup_\gamma \quad \mathbb{E}_\gamma[\tau - \gamma | \tau \geq \gamma]. \tag{5}$$

Problem 2:

$$\begin{aligned} \underset{\Pi}{\text{minimize}} \quad & \mathsf{CADD}(\Pi), \\ \text{subject to} \quad & \mathsf{FAR}(\Pi) \leq \alpha, \\ & \mathsf{PDC}_\ell(\Pi) \leq \beta_\ell, \quad \text{for} \quad \ell = 1, \cdots, L, \end{aligned} \tag{6}$$

where $0 \leq \alpha, \beta_\ell \leq 1$, $\ell = 1, \cdots, L$, are given constraints.

Note that when $\beta_\ell = 1$ $\forall \ell$ these problems reduce to the classical minimax formulations of [6] and [7], respectively.

In [8], an asymptotic lower bound on the CADD, and hence on the WADD, of any stopping time satisfying an FAR constraint of α is provided. See also [6] for WADD. Let

$$\Delta_\alpha = \{\Pi : \mathsf{FAR}(\Pi) \leq \alpha\}$$

denote the class of detection policies for which the FAR is no greater than the given number $0 < \alpha < 1$.

Theorem 2.1 (Asymptotic lower bound): Let the K-L numbers be positive and finite, $0 < D(f_{1,\ell} \,||\, f_{0,\ell}) < \infty$, $\ell = 1, \dots, L$. Then as $\alpha \to 0$,

$$\begin{aligned} \inf_{\Pi \in \Delta_\alpha} \mathsf{WADD}(\Pi) \geqslant & \inf_{\Pi \in \Delta_\alpha} \mathsf{CADD}(\Pi) \\ & \geq \frac{|\log \alpha|}{\sum_{\ell=1}^L D(f_{1,\ell} \,||\, f_{0,\ell})} (1 + o(1)). \end{aligned} \tag{7}$$

In the *centralized CuSum* algorithm, all the observations are used at the sensors ($M_{n,\ell} = 1 \; \forall n, \ell$), the raw observations are transmitted from the sensors to the fusion center at each time step ($Y_{n,\ell} = X_{n,\ell}$), and the CuSum algorithm (proposed in [9]) is applied to the vector observations received at the fusion center. That is, at the fusion center, the sequence $\{V_n\}$ is computed according to the following recursion: $V_0 = 0$, and for $n \geq 0$,

$$V_{n+1} = \max \left\{ 0,\ V_n + \sum_{\ell=1}^{L} \log L_\ell(X_{n+1,\ell}) \right\}, \qquad (8)$$

where $L_\ell(X) = \frac{f_{1,\ell}(X)}{f_{0,\ell}(X)}$ is the likelihood ratio in the ℓ-th sensor. A change is declared the first time V_n is above a threshold $A > 0$:

$$\tau_{\mathrm{c}} = \inf \{ n \geq 1 : V_n > A \}.$$

It is well known that the performance of the centralized CuSum algorithm is asymptotically equal to the lower bound provided in Theorem 2.1 [6]. However, $\mathsf{PDC}_\ell = 1\ \forall \ell$ for the centralized CuSum algorithm. Hence, the centralized CuSum algorithm is asymptotically optimal for Problem 1 and Problem 2 only when $\beta_\ell = 1$, $\forall \ell$.

Consider a policy in which, at each sensor every n^{th} sample is used, and raw observations are transmitted from each sensor to the fusion center, each time an observation is taken. At the fusion center, the CuSum algorithm, as defined above, is applied to the received samples. In this policy, the PDC_ℓ achieved is equal to $1/n$, $\forall \ell$. Using this scheme, any given constraints on the PDC_ℓ can be achieved by using every n^{th} sample, and by choosing a suitably large n. However, the detection delay for this scheme would be approximately n time that of the delay for the centralized CuSum algorithm, for the same false alarm rate.

Does there exist a policy using which any given constraints on the PDC_ℓ can be achieved, and yet the lower bound in Theorem 2.1 is achieved asymptotically? The answer to this question is in the affirmative. In the next section, we propose a data-efficient algorithm based on the DE-CuSum algorithm, which can be designed to satisfy FAR and PDC_ℓ constraints and attains the asymptotic lower bound in Theorem 2.1.

III. The DE-All algorithm

In this section, we propose the DE-All algorithm based on the DE-CuSum algorithm proposed in [5].

The CuSum statistic, computed using the observations at sensor ℓ, evolves recursively as follows: $C_{0,\ell} = 0$, and for $n \geq 0$,

$$C_{n+1,\ell} = \left(C_{n,\ell} + \log L_\ell(X_{n+1,\ell}) \right)^{+}. \qquad (9)$$

Thus, in the CuSum algorithm, the log likelihood ratio of the observations are accumulated over time. If the statistic $C_{n,\ell}$ goes below 0, it is reset to 0.

The Data-Efficient CuSum (DE-CuSum) algorithm is a modification of the CuSum algorithm that allows for adaptive on-off observation control. In the DE-CuSum algorithm, the DE-CuSum statistic evolves as follows. Start with $W_{0,\ell} = 0$, and for $n \geq 0$ use the following steps:

1) Take the first observation; fix $h_\ell \geq 0$ and $\mu_\ell > 0$.
2) If an observation is taken, then $W_{n,\ell}$ is updated using the recursion

$$W_{n+1,\ell} = \{ W_{n,\ell} + \log L_\ell(X_{n+1,\ell}) \}^{h_\ell +}, \qquad (10)$$

where $(x)^{h_\ell +} = \max\{x, -h_\ell\}$. Thus, the DE-CuSum statistic can become negative if $h_\ell > 0$.
3) If $W_{n,\ell} < 0$, then a sample is skipped and the statistic is updated using

$$W_{n+1,\ell} = \min\{ W_{n,\ell} + \mu_\ell, 0 \}. \qquad (11)$$

The evolution of the CuSum statistic and the DE-CuSum statistic, applied to the same set of samples, is plotted in Fig. 1.

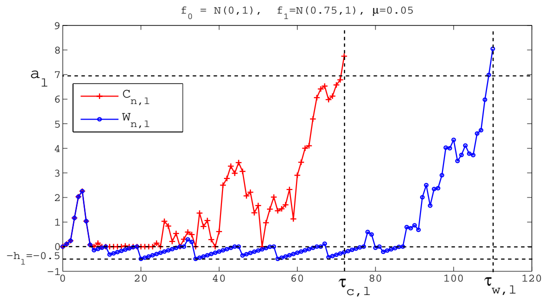

Fig. 1: Evolution of the CuSum statistic and the DE-CuSum statistic applied to the same set of samples.

If $h_\ell = \infty$, the evolution of the DE-CuSum statistic can be described as follows. The DE-CuSum statistic initially evolves according to the CuSum statistic till the statistic $W_{n,\ell}$ goes below 0. Once the statistic goes below 0, samples are skipped depending on the undershoot of $W_{n,\ell}$ (this is also the sum of the log likelihood ratios of the observations) and a pre-designed parameter μ_ℓ. Specifically, the statistic is incremented by μ_ℓ at each time step, and samples are skipped till $W_{n,\ell}$ goes above zero, at which time it is reset to zero. At this point, fresh observations are taken and the process is repeated on the fresh set of samples. If $h_\ell < \infty$, the maximum number of consecutive samples skipped is bounded by $h_\ell/\mu_\ell + 1$. This may be desired in some applications.

Before the change, the negative drift of the sum of the log likelihood ratios ensures that the DE-CuSum statistic goes below zero often. Also, each time the DE-CuSum statistic goes below 0, the positive increments of μ_ℓ ensures that the DE-CuSum statistic may go above 0, in which case new samples are taken. This ensures that the samples are taken and skipped regularly based on the 'state' of the system. Thus, the fraction of time observations are taken before change can be controlled using μ_ℓ. A smaller value of μ_ℓ would increase the average time spent by the DE-CuSum statistic below 0, thus decreasing the fraction of time observations are taken before change.

The DE-CuSum statistic can also be used to detect the change. This is because after the change and after a finite number of time steps, due to positive mean of the log likelihood ratio, the DE-CuSum statistic grows to ∞. In the DE-CuSum algorithm described in [5], the DE-CuSum statistic is computed using the above mentioned technique and a change is declared when the DE-CuSum statistic crosses a threshold. The properties of the DE-CuSum algorithm discussed above are used to establish certain optimality properties for the algorithm. Results that are of particular interest to this paper are summarized below. It is shown in [5] that

$$C_{n,\ell} \geq W_{n,\ell} \qquad \forall n \geq 0. \tag{12}$$

See Fig. 1. Define by

$$\begin{aligned} \tau_{\text{w},\ell} &= \inf\left\{n \geq 1 : W_{n,\ell} > a_\ell\right\}, \\ \tau_{\text{C},\ell} &= \inf\left\{n \geq 1 : C_{n,\ell} > a_\ell\right\} \end{aligned} \tag{13}$$

the corresponding stopping times. Thus, $\mathsf{FAR}(\tau_{\text{w},\ell}) \leq \mathsf{FAR}(\tau_{\text{C},\ell})$. Also it is shown that

$$\mathsf{CADD}(\tau_{\text{w},\ell}) \leq \mathsf{CADD}(\tau_{\text{C},\ell}) + \text{ constant}, \tag{14}$$

which implies that the centralized DE-CuSum algorithm attains asymptotic lower bound (7).

A. *The* DE − All *algorithm*

We now propose a decentralized algorithm, the DE-All algorithm, in which the DE-CuSum statistic $\{W_{n,\ell}\}_{n\geq 0}$ is computed at sensor ℓ using the local observation sequence $\{X_{n,\ell}\}_{n\geq 0}$ based on recursions (10) and (11). A '1' is transmitted each time the DE-CuSum statistic at any sensor is above a threshold. A change is declared the first time a '1' is received at the fusion center from *all* the sensors at the same time.

Mathematically, the DE-All algorithm can be written as follows. Let $d_\ell = D(f_{1,\ell} \,||\, f_{0,\ell}) / \sum_{k=1}^{L} D(f_{1,k} \,||\, f_{0,k})$.

*Algorithm 1 (*DE − All*: Π_{All}):* Start with $W_{0,\ell} = 0 \; \forall \ell$. Fix $\mu_\ell > 0$, $h_\ell \geq 0$, and $A > 0$. For $n \geq 0$ use the following control:

1) Update the DE-CuSum statistic at sensor ℓ according to the recursions (10) and (11).
2) Transmit $Y_{n,\ell} = 1$ only if $W_{n,\ell} > d_\ell A$.
3) At the fusion center stop at

$$\tau_{\text{DE-All}} = \inf\{n \geq 1 : Y_{n,\ell} = 1 \text{ for } \ell = 1, \cdots, L\}.$$

Remark 1: The parameters μ_ℓ and h_ℓ are selected to meet the constraint on the PDC_ℓ at each sensor, and the threshold A is selected to meet the constraint on the FAR.

B. *Asymptotic optimality of the* DE − All *algorithm*

We now show that the DE-All is asymptotically minimax optimal for both Problems 1 and Problem 2, for each fixed $\{\beta_\ell\}$, as $\alpha \to 0$.

Theorem 3.1 (Asymptotic optimality): Let $0 < \mathbb{E}_{1,\ell}[|\log L(X_{1,\ell})|]^3 < \infty$, $\ell = 1, \ldots, L$.

1) For any $\{\mu_\ell\}$ and $\{h_\ell\}$, setting $A = |\log \alpha|$ ensures that as $\alpha \to 0$, $\mathsf{FAR}(\Pi_{\text{All}}) \leq \alpha(1 + o(1))$.
2) There exists values of $\{\mu_\ell\}$, $\{h_\ell\}$, for which $\mathsf{PDC}_\ell(\Pi_{\text{All}}) \leq \beta_\ell$ for all $\ell = 1, \ldots, L$ and $A > 0$;
3) Let $\{\mu_\ell\}$ and $\{h_\ell\}$ be fixed numbers such that the constraints $\{\beta_\ell\}$ are met independently of the choice of A in the DE-All algorithm. If $h_\ell < \infty$ and $A = |\log \alpha|$, then as $\alpha \to 0$,

$$\mathsf{WADD}(\Pi_{\text{All}}) \sim \mathsf{CADD}(\Pi_{\text{All}}) \sim \frac{|\log \alpha|}{\sum_{\ell=1}^{L} D(f_{1,\ell} \,||\, f_{0,\ell})}.$$

Hence, the DE-All algorithm is asymptotically optimal in both Problem 1 and Problem 2 for each $\{\beta_\ell\}$ as $\alpha \to 0$.

Proof: The proof of this theorem is involved and cannot be presented due to the page limitation. It can be built based on [3, Theorem 3], [5, Theorem 5.2] and an r-quick convergence argument similar to that used in [10]. ■

Remark 2: The following approximation

$$\mathsf{PDC}_\ell \approx \frac{\mu_\ell}{\mu_\ell + D(f_{0,\ell} \,||\, f_{1,\ell})} \tag{15}$$

can be used to choose the value of μ_ℓ for a large h_ℓ to satisfy the PDC_ℓ constraint at each sensor (see [5] for a rigorous justification).

IV. Two more data-efficient algorithms for sensor networks

In the DE-All algorithm information is transmitted very rarely to the fusion center and in the binary form of '1's and '0's. This essentially means that the decision on the change is made at the sensors, since the change is declared at the fusion center the first time the change is 'sensed' by all the sensors simultaneously. This results in a poor performance.

In this section, we discuss two more data-efficient algorithms for sensor networks, initially proposed in [11]. The first one is also a distributed algorithm that we refer to as the *DE-Dist algorithm*. In this algorithm more information is transmitted from the sensors to the fusion center, as compared to that transmitted in the DE-All algorithm. This fact is utilized at the fusion center where an improved fusion technique (in fact the CuSum algorithm) is employed. The second algorithm discussed below is the *Serialized-DE-CuSum algorithm.* It is a centralized control based algorithm in which the fusion center executes the observation control and also makes the stopping decision.

A. *The* DE − Dist *algorithm*

In the DE-Dist algorithm, the DE-CuSum algorithm is used at each sensor for observation control. If an observation is taken at a sensor, then the observation is transmitted to the fusion center. At the fusion center, the CuSum algorithm is applied to the information received from the sensors to detect the change. Mathematically, the DE-Dist algorithm can be described as follows.

*Algorithm 2 (*DE − Dist*: Π_F):* Start with $W_{0,\ell} = 0 \; \forall \ell$. Fix $\mu_\ell > 0$, $h_\ell \geq 0$ and $A \geq 0$. For $n \geq 0$:

1) Update the DE-CuSum statistic at sensor ℓ according to the recursions (10) and (11).
2) Transmit $Y_{n,\ell} = \log L_\ell(X_{n,\ell})$ if $M_{n,\ell} = 1$.

3) At the fusion center compute the statistics $\{F_k\}$ using the CuSum recursion: $F_0 = 0$ and for $n \geq 0$,

$$F_{n+1} = \left(F_n + \sum_{\ell=1}^{L} Y_{n+1,\ell} \right)^{+}. \tag{16}$$

Stop and declare change at

$$\tau_{\text{DE-Dist}} = \inf\{n \geq 1 : F_n > A\}. \tag{17}$$

As in the DE-All algorithm, the parameters $\{\mu_\ell\}$, $\{h_\ell\}$ can be used to meet the constraint on the PDC_ℓ (e.g., (15) can be used). We conjecture that the DE-Dist algorithm is asymptotically optimal for both Problem 1 and Problem 2.

B. *The* Serialized – DE – CuSum *algorithm*

In the Serialized-DE-CuSum algorithm, the observation control is implemented by serializing the observations from the sensors. That is, the sequence $\{X_{1,1}, \cdots, X_{1,L}, X_{2,1}, \cdots, X_{2,L}, \ldots\}$ is considered, and the DE-CuSum algorithm (with a fixed μ and h) is applied to this serialized sequence. If $W_n(\ell)$ is the DE-CuSum statistic computed using the above serialized observations sequence, then a change is declared at

$$\tau_S = \inf\{n \geq 1 : W_n(L) > H\}.$$

If $f_{0,\ell} = f_0$ and $f_{1,\ell} = f_1$, i.e., the pre- and post-change distributions are the same for all the sensors, then it can be shown that setting $H = \log \frac{L}{\alpha}$ ensures that $\mathsf{FAR}(\tau_S) \leq \alpha$. Also the maximal expected delay is equal to the lower bound of Theorem 2.1. Thus, this algorithm is asymptotically optimal, as $\alpha \to 0$. However, note that since the observation control is executed by the fusion center, this algorithm is not a policy of type Π considered in Problem 1 and Problem 2. This algorithm will be essentially used as a benchmark for performance obtained in a data-efficient setting.

V. Numerical results

In Fig. 2 we compare operating characteristics (CADD vs. FAR) of the DE-All algorithm with that of the DE-Dist algorithm and the Serialized-DE-CuSum algorithm. The parameters used in the simulations are: $L = 10$, $f_{0,\ell} = \mathcal{N}(0,1)$, $\forall \ell$, and $f_{1,\ell} = \mathcal{N}(0.2,1)$, $\forall \ell$. We see that the DE-Dist algorithm performs better than the DE-All algorithm, as expected, since it transmits more information and uses a better fusion technique. Also, the Serialized-DE-CuSum algorithm performs better than both the DE-All algorithm and the DE-Dist algorithm, also as expected, since in the Serialized-DE-CuSum the observation control is executed by the fusion center. Also see [11].

VI. Conclusions

We have proposed minimax problem formulations for data-efficient quickest change detection in a decentralized system, e.g., a sensor network. Two data-efficient (DE) changepoint detection algorithms have been proposed. The first algorithm, the DE-All algorithm, can be efficiently used in a sensor network when there is a severe constraint on the communication between the sensors and the fusion center. We showed that this algorithm is asymptotically first-order optimal in the minimax problems with constraints imposed on both the FAR and the cost of taking observations in the pre-change mode. When the communication constraint is not so severe, we proposed to use another algorithm, the DE-Dist algorithm, in which more information is transmitted from the sensors to the fusion center and the fusion center uses a better fusion technique. We showed via simulations that the DE-Dist algorithm performs significantly better than the DE-All algorithm. We conjecture that this algorithm is also asymptotically minimax optimal, and a proof of this conjecture will be presented elsewhere.

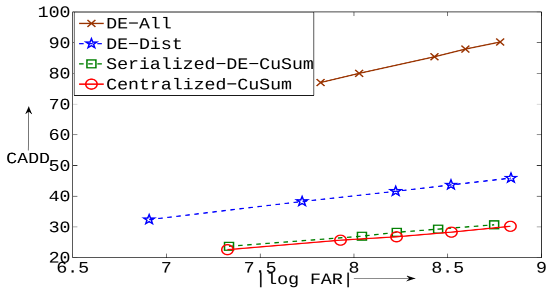

Fig. 2: Operating characteristics of the algorithms studied: $h_\ell = 10$, and μ_ℓ is selected to satisfy the PDC constraint of 0.5.

VII. Acknowledgments

This research was supported in part by the National Science Foundation (NSF) under grants CCF 08-30169, CCF 11-11342 and DMS 12-22498 and by the Defense Threat Reduction Agency (DTRA) under subcontract 147755 at the University of Illinois, Urbana-Champaign from prime award HDTRA1-10-1-0086, as well as by the Air Force Office of Scientific Research under MURI grant FA9550-10-1-0569, by the DTRA under grant HDTRA1-10-1-0086, by the Defense Advanced Research Projects Agency under grant W911NF-12-1-0034 and by the NSF under grants CCF-0830419 and DMS-1221888 at the University of Southern California, Department of Mathematics.

References

[1] V. V. Veeravalli and T. Banerjee, *Quickest Change Detection*. Elsevier: E-reference Signal Processing, 2013. http://arxiv.org/abs/1210.5552.

[2] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1657–1665, May 2001.

[3] Y. Mei, "Information bounds and quickest change detection in decentralized decision systems," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2669 –2681, July 2005.

[4] A. G. Tartakovsky and V. V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequential Analysis*, vol. 27, pp. 441–475, Oct. 2008.

[5] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in minimax settings," *ArXiv e-prints*, Nov. 2012. http://arxiv.org/abs/1211.3729.

[6] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, Dec. 1971.

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

[8] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2917 –2929, Nov. 1998.

[9] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, June 1954.

[10] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *SIAM Theory of Prob. and App.*, vol. 49, pp. 458–497, Sept. 2005.

[11] T. Banerjee and V. V. Veeravalli, "Energy-efficient quickest change detection in sensor networks," in *IEEE Statistical Signal Processing Workshop*, Aug. 2012.