DATA-EFFICIENT QUICKEST CHANGE DETECTION IN DISTRIBUTED AND MULTI-CHANNEL SYSTEMS

Taposh Banerjee and Venugopal V. Veeravalli

Dept of Electrical and Computer Engineering
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

ABSTRACT

A distributed or multi-channel system consisting of multiple sensors is considered. At each sensor a sequence of observations is taken, and at each time step, a summary of available information is sent to a central decision maker, called the fusion center. At some point of time, the distribution of observations at an unknown subset of the sensor nodes changes. The objective is to detect this change as quickly as possible, subject to constraints on the false alarm rate, the cost of observations taken at the sensors and the cost of communication between the sensors and the fusion center. Minimax formulations are proposed for this problem. An algorithm called DE-Censor-Sum is proposed, and is shown to be asymptotically optimal for the proposed formulations, for each possible post-change scenario, as the false alarm rate goes to zero. It is also shown, via numerical studies, that the DE-Censor-Sum algorithm performs significantly better than the approach of fractional sampling, where the cost constraints are met based on the outcome of a sequence of biased coin tosses, independent of the observation process.

Index Terms— Quickest change detection, observation control, transmission control, minimax, multi-channel systems, asymptotic optimality.

1 Introduction

In this paper we consider a distributed system which comprises of a set of sensors and a central decision maker called the fusion center. At each sensor a sequence of random variables is observed over time. At each time step, a processed version of the observations is transmitted from each sensor to the fusion center. At some point of time, called the change point, the distribution of the observations at some *unknown* subset of the sensors changes. The observations are independent conditioned on the change point, and identically distributed before and after the change point (this is called the i.i.d. model in the following). The distributions of the observations can vary from sensor to sensor. The objective is to detect this change as quickly as possible. Detecting the change even before the change actually occurs is considered as a false alarm and is not desirable. Hence, the change has to be detected subject to a constraint on the false alarm rate. In many practical systems where the above model is applicable, e.g., quality control, infrastructure, environment or habitat monitoring, and spectrum sensing in cognitive radios, there is a cost associated with acquiring observations at each sensor. Also, there is a cost associated with the communication between the sensors and the fusion center. That is there is a cost associated with acquiring data in the system. Thus, the change also has to be detected in a *data-efficient* way.

The classical centralized version of the quickest change detection problem, where a change has to detected in the distribution of a single sequence of random variables, is well studied in the literature [1], [2], [3], [4], [5]. See [6], [7], and [8] for some recent surveys. The objective in the classical setting is to detect a change in the distribution of random variables, so as to minimize some version of the average delay, subject to a constraint on the false alarm rate. Depending on the availability of the information on the distribution of the change point, the quickest change detection problem is either studied in the Bayesian setting of [1], or in non-Bayesian or minimax settings of [2] and [3].

In [9] and [10] we extended the classical quickest change detection formulations studied in [1], [2] and [3], by putting an additional constraint on the cost of observations used in the detection process. We proposed problem formulations, for the Bayesian setting in [9], and for two minimax settings in [10], in which the objective was to minimize some version of the average delay, subject to constraints on the false alarm rate and a version of the average number of observations taken before the change point. For the i.i.d. model, we proposed two-threshold extensions of the classical single-threshold algorithms, and showed that they are asymptotically optimal for the proposed formulations. We also showed via numerical results that the two-threshold algorithms we proposed provides significant gain in performance as compared to the approach of fractional sampling, in which the constraint on the observation cost is met by skipping samples randomly.

In [11], we used the insights obtained from [9] and [10] to propose data-efficient algorithms for distributed systems when the change affects all the sensors. We showed that the proposed algorithms have good trade-off curves and provide a significant gain in performance as compared to the approach of fractional sampling.

In this paper we extend the results from [9], [10] and [11] to study data-efficient quickest change detection (DE-QCD) in a distributed system described above, when the subset of sensors affected by the change is not known at the fusion center. Since, the knowledge of the distribution of the change point is generally not available in practice, we study the problem in minimax settings. Specifically, we propose minimax formulations for distributed systems in which the objective is to find a stopping time on the information received at the fusion center, so as to minimize a version of the worst case average delay. This delay has to be minimized subject to constraints

This research was supported in part by the National Science Foundation under grant CCF 08-30169, CCF 11-11342 and DMS 12-22498, through the University of Illinois at Urbana-Champaign. This research was also supported in part by the U.S. Defense Threat Reduction Agency through subcontract 147755 at the University of Illinois from prime award HDTRA1-10-1-0086.

on the false alarm rate, the cost of observations used at each sensor, and the cost of communication between each sensor and the fusion center. We propose an algorithm called the DE-Censor-Sum algorithm, and show that the algorithm is asymptotically optimal for the proposed formulations, for each possible post change scenario, as the false alarm rate goes to zero. We also show that the algorithm provides a significant gain in performance as compared to the approach of fractional sampling, where the constraints on the cost of data is met by skipping and transmitting samples randomly.

In the absence of a constraint on the cost of communication between the sensors and the fusion center, the distributed setting is referred to as the multi-channel setting. Since, the multi-channel setting is a special case of the distributed setting, the above results on the DE-Censor-Sum algorithm are valid for the multi-channel setting as well.

Bayesian quickest change detection in a sensor network with observation control is studied in [12], and with communication constraint is studied in [13] and [14]. However, the quickest change detection problem for sensor networks, when the subset of affected sensors is not known, under additional constraints on the observation cost and the cost of communication, and in a minimax setting, is not studied in the literature. Also, to the best of our knowledge, data-efficient quickest change detection in a multi-channel setting is not studied in the literature.

2 Centralized Minimax Formulations for DE-QCD and the DE-CuSum algorithm

Since, the formulations and algorithm proposed in this paper crucially depend on the formulations and the algorithm proposed in [10], in this section we provide a detailed overview of the results from [10]. In the following, we use $\mathbb{E}_n$, $\mathbb{P}_n$ to denote the expectation and the probability measure when the change occurs at time n, $n \leq \infty$. 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 $D(f \parallel g)$ to represent the K-L divergence between the 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.

In [10] we considered data-efficient quickest change detection in a single observation sequence. We considered an observation sequence $\{X_n\}$: $\{X_n\}$ are i.i.d. with probability density function (p.d.f.) f_0 before the change point γ, and are i.i.d. with p.d.f. f_1 after the change point γ. A decision maker observes the random variables $\{X_n\}$ over time and has to detect this change in distribution as quickly as possible, subject to constraints on the false alarm rate and the fraction of time observations are taken before change. Let M_n be the indicator random variable such that $M_n = 1$ if X_n is used for decision making, and $M_n = 0$ otherwise. Let

$$I_n = \left[M_1, \ldots, M_n, X_1^{(M_1)}, \ldots, X_n^{(M_n)}\right],$$

represent the information at time n. Here, $X_i^{(M_i)}$ represents X_i if $M_i = 1$, otherwise X_i is absent from the information vector I_n. Let τ be a stopping time on the information sequence $\{I_n\}$, that is $\mathbb{I}_{\{\tau=n\}}$ is a measurable function of I_n. Here, $\mathbb{I}_F$ represents the indicator of the event F. For time $n \geq 1$, based on the information vector I_n, a decision is made whether to *stop and declare change* ($\tau = n$) or *to continue taking observations* ($\tau > n$). If the decision is to *continue*, a decision is made as to whether to *take* or *skip* the observation at time $n+1$. Thus, M_{n+1} is a function of the information available at time n, i.e.,

$$M_{n+1} = \phi_n(I_n),$$

where, ϕ_n is the control law at time n. The decision of whether or not to take the first observation is taken without observing $\{X_n\}$. In the absence of a prior information on the distribution of γ, M_1 is typically set to 1, that is the first sample is always taken. A policy for data-efficient quickest change detection is $\Psi = \{\tau, \phi_0, \ldots, \phi_{\tau-1}\}$.

To capture the cost of observations used before γ, we proposed a new metric for data-efficiency in mimimax settings: the Pre-change Duty Cycle (PDC):

$$\mathsf{PDC}(\Psi) = \limsup_n \frac{1}{n}\mathbb{E}_n\left[\sum_{k=1}^{n-1} M_k \Big| \tau \geq n\right]. \tag{1}$$

We note that $\mathsf{PDC} \leq 1$. If in a policy all the samples are taken, then the PDC for that policy is 1. If every other sample is skipped, then the PDC for that policy is 0.5.

For delay and false alarm, we considered the metrics used in [2]: the Worst case Average Detection Delay (WADD)

$$\mathsf{WADD}(\Psi) = \sup_n \text{ess sup } \mathbb{E}_n\left[(\tau - n)^+ | I_{n-1}\right], \tag{2}$$

and the False Alarm Rate (FAR)

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

We considered the following data-efficient minimax formulation

Problem 1 ([10]).

$$\begin{aligned} &\underset{\Psi}{\textit{minimize}} && \mathsf{WADD}(\Psi), \\ &\textit{subject to} && \mathsf{FAR}(\Psi) \leq \alpha, \\ &\textit{and} && \mathsf{PDC}(\Psi) \leq \beta. \end{aligned} \tag{4}$$

Here, $0 \leq \alpha, \beta \leq 1$ *are given constraints.*

We also studied the data-efficient minimax formulation where instead of WADD, the following Conditional Average Detection Delay (CADD) metric from [3] was used:

$$\mathsf{CADD}(\Psi) = \sup_n \; \mathbb{E}_n[\tau - n | \tau \geq n]. \tag{5}$$

Problem 2 ([10]).

$$\begin{aligned} &\underset{\Psi}{\textit{minimize}} && \mathsf{CADD}(\Psi), \\ &\textit{subject to} && \mathsf{FAR}(\Psi) \leq \alpha, \\ &\textit{and} && \mathsf{PDC}(\Psi) \leq \beta. \end{aligned} \tag{6}$$

Here, $0 \leq \alpha, \beta \leq 1$ *are given constraints.*

We then proposed an algorithm, that we called the DE-CuSum algorithm, and showed that it is asymptotically optimal for both the above problems, as $\alpha \to 0$. The DE-CuSum algorithm is defined below.

Algorithm 1 ($\mathrm{DE-CuSum}$: $\Psi_{\mathrm{W}}(D, \mu, h)$)**.** *Start with* $W_0 = 0$ *and fix* $\mu > 0$, $D > 0$ *and* $h \geq 0$. *For* $n \geq 0$ *use the following control:*

$$M_{n+1} = \begin{cases} 0 & \textit{if } W_n < 0 \\ 1 & \textit{if } W_n \geq 0 \end{cases}$$

$$\tau_{\mathrm{W}} = \inf\{n \geq 1 : W_n > D\}.$$

The statistic W_n *is updated using the following recursions:*

$$W_{n+1} = \begin{cases} \min\{W_n + \mu, 0\} & \textit{if } M_{n+1} = 0 \\ (W_n + \log L(X_{n+1}))^{h+} & \textit{if } M_{n+1} = 1 \end{cases}$$

where $(x)^{h+} = \max\{x, -h\}$, *and* $L(X) = \frac{f_1(X)}{f_0(X)}$.

If $h = 0$, the DE-CuSum statistic W_n never becomes negative and hence reduces to the CuSum statistic [15] and evolves as: $W_0 = 0$, and for $n \geq 0$,

$$W_{n+1} = \max\{0, W_n + \log L(X_{n+1})\}.$$

The evolution of the DE-CuSum algorithm is plotted in Fig. 1. If $h < \infty$, the DE-CuSum statistic evolves as follows. Initially the DE-CuSum statistic evolves according to the CuSum statistic till the statistic W_n goes below 0. Once the statistic goes below 0, samples are skipped depending on the undershoot of W_n (this is also the sum of the log likelihood ratio of the observations) and the design parameter μ. Specifically, the statistic is incremented by μ at each time step, and samples are skipped till W_n goes above zero, at which time it is reset to zero. At this point, fresh observations are taken and the process is repeated till the statistic crosses the threshold D, at which time a change is declared. The parameter μ is a substitute for the Bayesian prior ρ that is used in the DE-Shiryaev algorithm described in [9], and is chosen to meet the constraint on the PDC. Thus, the DE-CuSum algorithm is a sequence of SPRTs ([16], [17]) intercepted by 'sleep' times controlled by the undershoot and the parameter μ. If $h < \infty$, the number of consecutive samples skipped is bounded by $h/\mu + 1$.

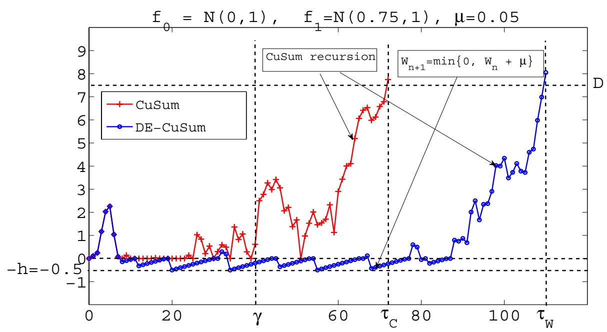

Fig. 1*:* Typical evolution of the CuSum statistic and the DE-CuSum statistic evaluated using the same set of samples. Note that the CuSum statistic is always greater than the DE-CuSum statistic.

Let τ_C represent the stopping time for the CuSum algorithm. The following result is proved in [10].

Theorem 2.1 ([10])**.** *For any μ, h, and D,*

$$\begin{aligned} \mathsf{FAR}(\Psi_W) &\leq \mathsf{FAR}(\tau_C), \\ \mathsf{WADD}(\Psi_W) &\leq \mathsf{WADD}(\tau_C) + \textit{ constant }. \end{aligned} \tag{7}$$

Also, there exists μ^* and h^* such that

$$\mathsf{PDC}(\Psi_W(D, \mu^*, h^*)) \leq \beta, \quad \forall D \geq 0.$$

Remark 1. *Thus, setting $D = |\log \alpha|$ ensures that $\mathsf{FAR}(\Psi_W) \leq \alpha$ [2]. It is well known that the* CuSum *algorithm is asympototically optimal for both Problem 1 and Problem 2 when $\beta = 1$, with delay $\mathsf{WADD}(\tau_C) \sim \frac{|\log \alpha|}{D(f_1 \,||\, f_0)}$, as $\alpha \to 0$ [18]. This implies that the* $\mathrm{DE-CuSum}$ *algorithm is asymptotic optimal, for both Problem 1 and Problem 2 for each fixed β, as $\alpha \to 0$, with the same asymptotic delay.*

A good approximation for PDC *for a large h is*

$$\mathsf{PDC} \approx \frac{\mu}{\mu + D(f_0 \,||\, f_1)}. \tag{8}$$

Thus, to design the $\mathrm{DE-CuSum}$ *algorithm to achieve a smaller value of* PDC*, that is to design the algorithm to drop a larger fraction of samples before change, one has to select a smaller value for the parameter μ.*

In [10], we also showed via simulations that the DE-CuSum algorithm provides a significant gain in performance as compared to the approach of fractional sampling, where the CuSum algorithm is used and the PDC constraint is met by skipping samples randomly.

3 Problem formulation for distributed systems and the DE-Censor-Sum algorithm

The distributed system is assumed to consist of L sensors and a fusion center. At sensor ℓ the sequence $\{X_{n,\ell}\}_{n \geq 1}$ is observed, where n is the time index. At γ, the distribution of $\{X_{n,\ell}\}$ in a subset $\{k_1, k_2, \cdots, k_m\} \subset \{1, 2, \cdots, L\}$ of the streams changes, from $f_{0,\ell}$ to say $f_{1,\ell}$. $\{X_{n,\ell}\}$ are independent across indices n and ℓ conditioned on γ. The distributions $f_{0,\ell}$ and $f_{1,\ell}$ are known to the decision maker but the affected subset $\{k_1, k_2, \cdots, k_m\}$ is not known.

In analogy with the notation used in Section 2, 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 $I_{n,\ell} = \left[M_{1,\ell}, \ldots, M_{n,\ell}, X_{1,\ell}^{(M_{1,\ell})}, \ldots, X_{n,\ell}^{(M_{n,\ell})}\right]$. 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. Thus, the decision to take or skip a sample at sensor ℓ, and the decision on what to transmit from sensor ℓ to the fusion center, is based on the information at sensor ℓ alone. Let

$$Y_n = \{Y_{n,1}, \cdots, Y_{n,L}\}$$

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

$$\Pi = \{\tau, \{\phi_{n,\ell}\}_{n=0}^{\tau-1}, \{g_{n,\ell}\}_{n=1}^{\tau}\}$$

be a policy for data-efficient quickest change detection in distributed systems.

Define PDC_ℓ, the PDC for sensor ℓ as

$$\mathsf{PDC}_\ell(\Pi) = \limsup_n \frac{1}{n} \mathbb{E}_n \left[\sum_{k=1}^{n-1} M_{k,\ell} \middle| \tau \geq n \right]. \tag{9}$$

Also define $T_{n,\ell}$ as $T_{n,\ell} = 1$ if $Y_{n,\ell} \neq$ NULL. We now define the new metric for communication efficiency, the Pre-change Transmission Cost at sensor ℓ (PTC_ℓ)

$$\mathsf{PTC}_\ell(\Pi) = \limsup_n \frac{1}{n} \mathbb{E}_n \left[\sum_{k=1}^{n-1} T_{k,\ell} \middle| \tau \geq n \right]. \tag{10}$$

If in a policy every sample is taken and some information is transmitted at every time slot at all the sensors, then for that policy $\mathsf{PDC}_\ell = \mathsf{PTC}_\ell = 1$, $\forall \ell$. If transmissions happen from the sensors only in every alternate time slots, then $\mathsf{PTC}_\ell = 0.5$, $\forall \ell$.

The objective here is to solve the following data-efficient extensions of Problem 1 and Problem 2.

Problem 3.

$$\begin{aligned} &\underset{\Pi}{\textit{minimize}} && \mathsf{WADD}(\Pi), \\ &\textit{subject to} && \mathsf{FAR}(\Pi) \leq \alpha, \\ & && \mathsf{PDC}_\ell(\Pi) \leq \beta_\ell, \ \textit{ for } \ \ell = 1, \cdots, L, \\ &\textit{and} && \mathsf{PTC}_\ell(\Pi) \leq \sigma_\ell, \ \textit{ for } \ \ell = 1, \cdots, L. \end{aligned} \tag{11}$$

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

Problem 4.

$$
\begin{aligned}
\underset{\Pi}{\textit{minimize}} \quad & \mathsf{CADD}(\Pi), \\
\textit{subject to} \quad & \mathsf{FAR}(\Pi) \leq \alpha, \\
& \mathsf{PDC}_\ell(\Pi) \leq \beta_\ell, \;\; \textit{for} \;\; \ell = 1, \cdots, L, \\
\textit{and} \quad & \mathsf{PTC}_\ell(\Pi) \leq \sigma_\ell, \;\; \textit{for} \;\; \ell = 1, \cdots, L.
\end{aligned} \tag{12}
$$

Here, $0 \leq \alpha, \beta_\ell, \sigma_\ell \leq 1$, *for* $\ell = 1, \cdots, L$, *are given constraints.*

Remark 2. *If* $\sigma_\ell = 1 \, \forall \ell$, *then Problem 3 and Problem 4 specializes to problems for multi-channel systems.*

We now propose the DE-Censor-Sum algorithm. In the DE-Censor-Sum algorithm, the DE-CuSum algorithm is used at each sensor (data-efficiency). If the DE-CuSum statistic is above a threshold, then the statistic is transmitted to the fusion center (censoring). A change is declared at the fusion center, if the sum of the transmitted statistics from all the sensors is larger than another threshold (thus detecting a change without knowing the subset of affected sensors). Let $d_\ell = \frac{D(f_{1,\ell} \,||\, f_{0,\ell})}{\sum_{k=1}^{L} D(f_{1,k} \,||\, f_{0,k})}$, and let $W_{n,\ell}$ be the DE-CuSum statistic at sensor ℓ.

Algorithm 2 (DE − Censor − Sum: $\Pi_{\mathrm{W}}(\{\mu_\ell\}, \{h_\ell\}, D, A)$)**.** *Start with* $W_{0,\ell} = 0 \; \forall \ell$. *Fix* $\mu_\ell > 0$, $h_\ell \geq 0$, $D \geq 0$ *and* $A \geq 0$. *For* $n \geq 0$ *use the following control:*

1. *Use the policy* $\Psi_{\mathrm{W}}(\infty, \mu_\ell, h_\ell)$ *at sensor* ℓ
2. *Transmit* $\;Y_{n,\ell} = W_{n,\ell} \;$ *if* $\; W_{n,\ell} > d_\ell D$
3. *Stop at*
$$\tau_{\textit{DE-Censor-Sum}} = \inf\{n \geq 1 : \sum_{\ell=1}^{L} Y_{n,\ell} > A\}.$$

In the DE-Censor-Sum algorithm, the parameters μ_ℓ and h_ℓ are chosen to control the PDC_ℓ at each sensor (see approximation (8)). Choosing a large value for the thresholds D and A leads to fewer transmissions (lower PTC_ℓ) and smaller FAR, respectively.

With $D = 0$ and $h_\ell = 0$, $\forall \ell$, the DE-CuSum algorithm at each sensor reduces to the CuSum algorithm, and $Y_{n,\ell} = W_{n,\ell} \; \forall n, \ell$. In this case, the DE-Censor-Sum algorithm reduces to the N_{sum} algorithm proposed in [19] (also see [20]). Note that for the N_{sum} algorithm, $\mathsf{PDC}_\ell = \mathsf{PTC}_\ell = 1 \; \forall \ell$.

We now prove the asymptotic optimality of the DE-Censor-Sum algorithm. Let CADD_κ and WADD_κ be the CADD and WADD evaluated with respect to the probability measure under which the subset κ is affected at the change point.

Theorem 3.1. *For any* $\{\mu_\ell\}$, $\{h_\ell\}$ *and D, setting*

$$A = A_\alpha = |\log \alpha| + (L - 1 + o(1)) \log |\log \alpha|$$

ensures that $\mathsf{FAR}(\Pi_W) \leq \alpha(1 + o(1))$ *as* $\alpha \to 0$. *There exists* $\{\mu_\ell^*\}$, $\{h_\ell^*\}$ *and* D^* *such that*

$$
\begin{aligned}
\mathsf{PDC}_\ell\big(\Pi_W(\{\mu_\ell^*\}, \{h_\ell^*\}, D, A)\big) &\leq \beta_\ell, \; \forall \ell, \; \forall D, A, \\
\mathsf{PTC}_\ell\big(\Pi_W(\{\mu_\ell\}, \{h_\ell\}, D^*, A)\big) &\leq \sigma_\ell, \; \forall \ell, \; \forall \{\mu_\ell\}, \{h_\ell\}, A.
\end{aligned} \tag{13}
$$

Finally if $A = A_\alpha$ *and* $h_\ell < \infty \, \forall \ell$, *then for each fixed* $\{\mu_\ell\}$, $\{h_\ell\}$ *and D, and for each possible subset* κ, *as* $\alpha \to 0$,

$$\mathsf{CADD}_\kappa(\Pi_W) \sim \mathsf{WADD}_\kappa(\Pi_W) \sim \frac{|\log \alpha|}{\sum_{i=1}^{m} D(f_{1,k_i} \,||\, f_{0,k_i})}.$$

Remark 3. *Since, the right hand side is the best possible asymptotic delay for an* FAR *constraint of* α *[18], the* DE − Censor − Sum *algorithm is asymptotically optimal for both Problem 3 and Problem 4, for each fixed* $\{\beta_\ell\}$, $\{\sigma_\ell\}$, *as* $\alpha \to 0$. *Also, since the algorithm is asymptotically optimal for each fixed* $\{\sigma_\ell\}$, *it is also asymptotically optimal for the multi-channel setting* ($\sigma_\ell = 1 \; \forall \ell$).

4 Numerical results

We now compare the performance of the DE-Censor-Sum algorithm with the fractional sampling scheme for $L = 10$, $f_{0,\ell} = \mathcal{N}(0, 1)$, $f_{1,\ell} = \mathcal{N}(0.5, 1)$, and for the PDC and PTC constraints of $\beta_\ell = \sigma_\ell = 0.5 \, \forall \ell$. We consider two different post-change scenarios: $m = 2$ and $m = 6$. Recall that m is the number of sensors affected by the change at the change point. We restrict our numerical study to the comparison of the CADD performance. Similar comparison can be obtained for the WADD as well.

In the fractional sampling scheme, the CuSum algorithm is used at each sensor, and samples are skipped based on the outcome of a sequence of fair coin tosses, independent of the observation process. If an observation is taken at a sensor, the CuSum statistic is transmitted to the fusion center. Thus, achieving the constraints on the PDC and PTC. At the fusion center a change is declared the first time the sum of the CuSum statistics from all the sensors crosses a threshold. At the fusion center, in the absence of any transmission from a sensor, its CuSum statistics from the last transmission is used to compute the sum. For the DE-Censor-Sum algorithm, we set $D = 0$, $\{h_\ell = h = 10\} \; \forall \ell$, and use the approximation (8) to select μ_ℓ. This ensures that the PDC and PTC constraints are satisfied for the DE-Censor-Sum algorithm. We also compare these two schemes with the N_{sum} algorithm of [19] which as discussed in the previous section has $\mathsf{PDC}_\ell = \mathsf{PTC}_\ell = 1$, $\forall \ell$. The comparison is plotted in Fig. 2.

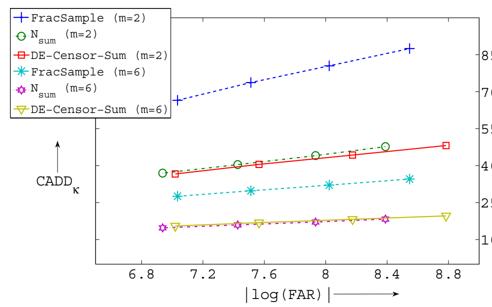

Fig. 2*:* Comparison of the CADD performances of the DE-Censor-Sum algorithm, the N_{sum} algorithm, and the fractional sampling scheme. Parameters used are $L = 10$, $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.5, 1)$, with PDC and PTC constraints of 0.5. m is the number of affected sensors post-change.

The figure shows that the DE-Censor-Sum algorithm provides a significant gain in performance as compared to the approach of fractional sampling. Also, when $m = 6$, even after dropping 50% of the samples at each sensor (and hence also transmitting only 50% of the time) there is negligible loss in performance, as compared to sensing and transmitting all the time (i.e., using N_{sum}). For $m = 2$ the figure shows that there is in fact a gain in performance by 'sleeping' 50% of the time. This is consistent with the observation in [20], where it was observed that censoring suppresses the noise from sensors not affected by change, and hence provides better false alarm-delay trade-off.

5 Conclusions

In this paper we studied data-efficient quickest change detection in multi-channel and distributed systems where the subset of sensors affected by the change is unknown to the decision maker. We proposed minimax formulations for the problem and proposed a data-efficient algorithm called the DE-Censor-Sum algorithm. We showed that the algorithm is asymptotically optimal for the proposed formulations, and performs significantly better than the approach of fractional sampling, where the constraints on the cost of data is met by skipping samples and transmitting randomly.

6 References

[1] A. N. Shiryaev, “On optimum methods in quickest detection problems,” *Theory of Prob and App.*, vol. 8, pp. 22–46, 1963.

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

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

[4] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *Ann. Statist.*, vol. 14, pp. 1379–1387, Dec. 1986.

[5] A. G. Tartakovsky, M. Pollak, and A. Polunchenko, “Third-order asymptotic optimality of the generalized Shiryaev-Roberts changepoint detection procedures,” *ArXiv e-prints*, May 2010.

[6] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press, 2009.

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

[8] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. Statistics, CRC Press, 2013.

[9] T. Banerjee and V. V. Veeravalli, “Data-efficient quickest change detection with on-off observation control,” *Sequential Analysis*, vol. 31, pp. 40–77, Feb. 2012.

[10] 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`.

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

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

[13] L. Zacharias and R. Sundaresan, “Decentralized sequential change detection using physical layer fusion,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 4999–5008, Dec. 2008.

[14] T. Banerjee, V. Sharma, V. Kavitha, and A. K. JayaPrakasam, “Generalized analysis of a distributed energy efficient algorithm for change detection,” *IEEE Trans. Wireless Commun.*, vol. 10, pp. 91–101, Jan. 2011.

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

[16] A. Wald and J. Wolfowitz, “Optimum character of the sequential probability ratio test,” *Ann. Math. Statist.*, vol. 19, no. 3, pp. pp. 326–339, 1948.

[17] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. Springer series in statistics, Springer-Verlag, 1985.

[18] 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.

[19] Y. Mei, “Efficient scalable schemes for monitoring a large number of data streams,” *Biometrika*, vol. 97, pp. 419–433, Apr. 2010.

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