A Computationally Efficient Algorithm for Quickest Change Detection in Anonymous Heterogeneous Sensor Networks

Zhongchang Sun*, Qunwei Li†, Ruizhi Zhang‡ , Shaofeng Zou*
* University at Buffalo, the State University of New York, Buffalo, NY, USA, {zhongcha, szou3}@buffalo.edu
† Ant Financial, China, qunwei.qw@antfin.com
‡ University of Nebraska-Lincoln, Lincoln, NE, USA, rzhang35@unl.edu

***Abstract*—The problem of quickest change detection in anonymous heterogeneous sensor networks is studied. The sensors are clustered into K groups, and different groups follow different data generating distributions. At some unknown time, an event occurs in the network and changes the data generating distribution of the sensors. The goal is to detect the change as quickly as possible, subject to false alarm constraints. The anonymous setting is studied, where at each time step, the fusion center receives unordered samples without knowing which sensor each sample comes from, and thus does not know its exact distribution. In [1], an optimal algorithm was provided, which however is not computational efficient for large networks. In this paper, a computationally efficient test is proposed and a novel theoretical characterization of its false alarm rate is further developed.**

I. Introduction

Suppose a network consists of n sensors and a fusion center. At some unknown time, an event occurs in the network, and causes a change in the data generating distribution of the sensors. The goal is to detect the change as quickly as possible subject to false alarm constraints. We consider a general setting with heterogeneous sensors, where the sensors can be clustered into K groups, and different groups follow different data generating distributions. In this paper, we investigate the scenario where the sensors are anonymous. Specifically, the fusion center does not know which sensor each sample comes from (see e.g., [2], [3] for anonymous data collection approaches). The anonymous and heterogeneous setting finds a wide range of applications in sensor networks in social settings [4], where human participants are involved, and thus privacy and anonymity are required to protect the participants.

In this paper, we investigate the anonymous setting. Existing approaches for quickest change detection (QCD) in sensor networks [5]–[14] are not applicable since the fusion center is not able to compute one CuSum statistic for each sensor.

In [15], the binary hypothesis testing problem under the anonymous setting was considered and an optimal mixture likelihood ratio test (MLRT) was developed. In [1], the QCD problem under the anonymous setting was investigated, and a mixture CuSum algorithm was constructed based on the MLRT, and was further shown to be exactly optimal under Lorden's criterion [16]. However, the computational complexity of mixture CuSum increases exponentially with n, which limits its applications in large networks. In this paper, we propose a computationally efficient test and further derive a lower bound on its worst-case average run length to false alarm, so that a threshold can be chosen analytically for false alarm control. We provide numerical results to demonstrate the performance of our test and its computational efficiency.

The work of Z. Sun and S. Zou is partially supported by a National Science Foundation (NSF) award under grant CCF-1948165.

II. Problem Formulation

Consider a network consisting of n sensors. The sensors are heterogeneous and can be divided into K groups. Each group k has n_k sensors, $1 \leq k \leq K$. Let $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_K]^T$, where $\alpha_k = \lim_{n\to\infty} \frac{n_k}{n}$. The distributions of the observations in group k are $p_{\theta,k}$, $\theta \in \{0, 1\}$. Let $\boldsymbol{P}_\theta = [p_{\theta,1} \cdots p_{\theta,K}]^T$. We assume that $\boldsymbol{\alpha}^T \boldsymbol{P}_0 \neq \boldsymbol{\alpha}^T \boldsymbol{P}_1$ almost everywhere.

The centralized setting is considered, where there is a fusion center. The sensors are anonymous, i.e., the fusion center doesn't know which group of sensors that each observation comes from. The fusion center only knows the distributions $p_{\theta,k}$, $\theta \in \{0, 1\}$ and the number of sensors n_k in each group k. We focus on discrete distributions, that is, the cardinality of $\mathcal{X}$ is finite, where $\mathcal{X}$ denotes the alphabet of distributions. Denote by $\mathcal{P}_\mathcal{X}$ the set of all distributions supported on $\mathcal{X}$.

In anonymous networks, unordered samples are observed sequentially. Let $X^n[t] = \{X_1[t], \ldots, X_n[t]\}$ be the n collected samples at time t, which are assumed to be independent. We further assume that $X^n[t_1]$ is independent from $X^n[t_2]$ for any $t_1 \neq t_2$. Denote by $\sigma_t(i) \in \{1, \ldots, K\}$ the label of the group that $X_i[t]$ comes from, i.e., $X_i[t] \sim p_{\theta,\sigma_t(i)}$. Due to the anonymity, $\sigma_t(i)$, $i = 1, \ldots, n$, are *unknown* to the fusion center. There are $\binom{n}{n_1,\ldots,n_K}$ possible $\sigma_t : \{1, \ldots, n\} \to \{1, \ldots, K\}$ satisfying $|\{i : \sigma_t(i) = k\}| = n_k, \forall k = 1, \ldots, K$. We denote the collection of all such labelings by $\mathcal{S}_{n,\lambda}$, where $\lambda = \{n_1, \ldots, n_K\}$.

At some unknown time ν, an event occurs and changes the data generating distributions of the sensors. Before the change, i.e., $t < \nu$, $X^n[t] \sim \mathbb{P}_{0,\sigma_t} \triangleq \prod_{i=1}^n p_{0,\sigma_t(i)}$, for some $\sigma_t \in \mathcal{S}_{n,\lambda}$. After the change, i.e., $t \geq \nu$, $X^n[t] \sim \mathbb{P}_{1,\sigma_t} \triangleq \prod_{i=1}^n p_{1,\sigma_t(i)}$, for some $\sigma_t \in \mathcal{S}_{n,\lambda}$. Note that σ_t may change with time, i.e., σ_{t_1} may not be the same as σ_{t_2}, for $t_1 \neq t_2$. Let $\mathbf{X}^n[t_1, t_2] = \{X^n[t_1], \cdots, X^n[t_2]\}$, for $t_1 \leq t_2$.

The objective is to detect the change at time ν as quickly as possible subject to false alarm constraints. In this paper, we consider a deterministic unknown change point ν, and we define the worst-case average detection delay (WADD) under Lorden's criterion [16] and worst-case average run length (WARL) for any stopping time τ as follows:

$$\begin{aligned} &\text{WADD}(\tau) \triangleq \sup_{\nu \geq 1} \sup_{\Omega} \text{ess}\sup \mathbb{E}_{\Omega}^{\nu}\left[(\tau-\nu)^{+}|\mathbf{X}^n[1,\nu-1]\right], \\ &\text{WARL}(\tau) \triangleq \inf_{\Omega} \mathbb{E}_{\Omega}^{\infty}[\tau]. \end{aligned} \tag{1}$$

where $\Omega = \{\sigma_1, \sigma_2, ..., \sigma_\infty\}$, $\mathbb{E}_{\Omega}^{\nu}$ denotes the expectation when the change is at ν, and samples are labeled by σ_t.

The goal is to design a stopping rule that minimizes the WADD subject to a constraint on the WARL:

$$\inf_{\tau:\text{WARL}(\tau)\geq\gamma} \text{WADD}(\tau). \tag{2}$$

III. Exactly Optimal Mixture CuSum

In [1], the following mixture CuSum algorithm was proposed for the QCD problem in (2):

$$T^*(b) = \inf\left\{t : \max_{1\leq k\leq t} \sum_{i=k}^{t} \log \frac{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n[i])}{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n[i])} \geq b\right\}, \tag{3}$$

where $b > 0$ is the threshold. This test was shown to be exactly optimal for the problem in (2). However, the mixture likelihood ratio $\frac{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n[i])}{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n[i])}$ needs to compute the average of the likelihood over all possible $\sigma \in \mathcal{S}_{n,\lambda}$. Note that the size of $\mathcal{S}_{n,\lambda}$ is $\binom{n}{n_1,\cdots,n_K}$. From the exponential bounds on the size of a type class [17], we have that $\frac{2^{nH\left([\frac{n_1}{n}\cdots\frac{n_K}{n}]\right)}}{(n+1)^{|\mathcal{X}|}} \leq \binom{n}{n_1,\cdots,n_K} \leq 2^{nH\left([\frac{n_1}{n}\cdots\frac{n_K}{n}]\right)}$, where $H\left([\frac{n_1}{n}\cdots\frac{n_K}{n}]\right)$ denotes the entropy of $[\frac{n_1}{n}\cdots\frac{n_K}{n}]$. As $n\to\infty$, we have that $\lim_{n\to\infty} H\left([\frac{n_1}{n}\cdots\frac{n_K}{n}]\right) = H(\boldsymbol{\alpha})$. Therefore, the computational complexity of mixture CuSum increases almost exponentially with n, which limits its practical applications in large networks. This motivates the need for computationally efficient tests for large networks.

IV. An Efficient Algorithm

In this section, we propose a computationally efficient algorithm. We derive a lower bound on its WARL so that a threshold can be chosen analytically for false alarm control. The WADD for the computationally efficient algorithm is challenging. We will then investigate the detection delay when the change occurs at $\nu = 1$.

A. Algorithm Construction

We first introduce some useful results that motivate the design of our algorithm. Let Π_{X^n} denote the empirical distribution of samples X^n, and $T(\Pi_{X^n})$ denote the type class of Π_{X^n} [17]. It can be firstly shown that [15]

$$\frac{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n)}{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n)} = \frac{\mathbb{P}_{1,\sigma}\left(T(\Pi_{X^n})\right)}{\mathbb{P}_{0,\sigma}\left(T(\Pi_{X^n})\right)}. \tag{4}$$

The right hand side of (4) is a function of the empirical distribution Π_{X^n}. Let $D(P\|Q)$ denote the Kullback-Leibler divergence between two distributions P and Q. Then the computation of likelihood ratio in mixture CuSum can be converted into an optimization problem when n is large [15]: for any $\theta \in \{0,1\}$ and any $\sigma \in \mathcal{S}_{n,\lambda}$,

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\theta,\sigma}\left(T(Q_n)\right) \\ &= -\inf_{\substack{\boldsymbol{U}=(U_1,\ldots,U_K)\in(\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T\boldsymbol{U}=Q}} \sum_{k=1}^{K} \alpha_k D(U_k||p_{\theta,k}), \end{aligned} \tag{5}$$

where $\mathcal{P}_n$ denote the set of types with denominator n, $Q_n \in \mathcal{P}_n$ denotes a sequence of distributions and $\lim_{n\to\infty} Q_n = Q$.

The right hand side of (5) is a convex optimization problem with linear constraints, which can be solved efficiently using standard optimization tools [18], [19]. Its computational complexity is independent of the number of sensors. Therefore, for large n, the mixture of the likelihood over σ in (4) can be approximated by convex optimization.

Let $\boldsymbol{P} = [P_1 \cdots P_K]^T$, where $P_k \in \mathcal{P}_{\mathcal{X}}$, $1 \leq k \leq K$. For any distribution $Q \in \mathcal{P}_{\mathcal{X}}$, define the following function of Q:

$$f_{\boldsymbol{P}}(\boldsymbol{\alpha}, Q) = \inf_{\substack{(U_1,\ldots,U_K)\in(\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T\boldsymbol{U}=Q}} \sum_{k=1}^{K} \alpha_k D(U_k||P_k). \tag{6}$$

Intuitively, an algorithm for problem in Section II can be constructed by approximating the mixture likelihood ratio at time t in the mixture CuSum algorithm using $f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[t]})$. However, the lower bound on WARL for this algorithm is difficult to derive due to the "inf" in the test statistic. In the following, we construct a novel test that can be updated recursively, and for which the lower bound on WARL can be theoretically characterized.

Let $\hat{\nu}_t$ denote the change point estimate at time t. Denote by $\hat{t} \triangleq t - \hat{\nu}_t + 1$. We then design our detection statistic to approximate $\sum_{i=k}^{t} \log \frac{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}(X^n[i])}{\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}(X^n[i])}$ in (3):

$$W[t] = \hat{t}n\left[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]})\right]. \tag{7}$$

Instead of using a maximum likelihood approach to estimate $\hat{\nu}_t$ as in (3), which is not computationally efficient here, since $\hat{\nu}_t$ also appears in $\mathbf{X}^n[\hat{\nu}_t, t]$, we design a recursive way of updating $\hat{\nu}_t$: let $\hat{\nu}_0 = 0$; if $W[t] \leq 0$, $\hat{\nu}_{t+1} = t+1$, and if $W[t] > 0$, $\hat{\nu}_{t+1} = \hat{\nu}_t$. Then, $\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}$ can also be updated recursively: if $W[t] \leq 0$, $\Pi_{\mathbf{X}^n[\hat{\nu}_{t+1},t+1]} = \Pi_{\mathbf{X}^n[t+1]}$, and if $W[t] > 0$, $\Pi_{\mathbf{X}^n[\hat{\nu}_{t+1},t+1]} = \frac{\hat{t}\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]} + \Pi_{\mathbf{X}^n[t+1]}}{t-\hat{\nu}_{t+1}+2}$.

We next provide a heuristic explanation of how $W[t]$ evolves in the pre- and post-change regimes. According to the Glivenko–Cantelli theorem [20], before the change point ν, as $n\to\infty$, $\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}$ converges to $\boldsymbol{\alpha}^T\boldsymbol{P}_0$ almost surely. It can be easily seen that $f_{\boldsymbol{P}}(\boldsymbol{\alpha}, Q) \geq 0$ for any $\boldsymbol{\alpha}, \boldsymbol{P}$ and Q. The equality holds if and only if $\boldsymbol{\alpha}^T\boldsymbol{P} = Q$ almost everywhere. This implies that $f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T\boldsymbol{P}_0) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T\boldsymbol{P}_0) < 0$. Therefore, before the change point ν, for large n, $W[t]$ has a negative drift. Similarly, after the change point ν, for large n, $W[t]$ has a positive drift and evolves towards ∞. This motivates us to construct the following computationally efficient test:

$$\tau_e = \inf\Big\{t \geq 1 : W[t] \geq b\Big\}. \tag{8}$$

The computation cost of τ_e mainly lies in the update of the empirical distribution and the optimization step. The computational complexity of updating the empirical distribution increases linearly with n, and the computational complexity of the optimization problem is independent of n. Therefore, the computationally efficient test is more efficient than the optimal mixture CuSum algorithm when n is large.

B. Lower Bound on WARL

In this section, we derive the WARL lower bound for our computationally efficient test in (8).

Theorem 1. *Define* $\Gamma \triangleq \big\{\mu \in \mathcal{P}_{\mathcal{X}} | f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \mu) > f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu)\big\}$. *Let* $h = \inf_{\substack{(U_1,\dots,U_K)\in(\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T\boldsymbol{U}\in\Gamma}} \sum_{k=1}^K n_k D(U_k||P_{0,k})$. *Then* $h > 0$, *and for any* Ω,

$$\mathbb{E}_{\Omega}^{\infty}\big[\tau_e(b)\big] \geq \frac{e^b}{\big(\frac{b}{h}+1\big)\big(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\big)}. \tag{9}$$

Proof. Let $Y = \inf\{t \geq 1 : W[t] \leq 0\}$ be the first regeneration time. For any Ω and $m \geq 1$, we have that

$$\begin{aligned}
&\mathbb{P}_{\Omega}^{\infty}(Y > m) = \mathbb{P}_{\Omega}^{\infty}\big(W[t] > 0, \forall t \in [1, m]\big) \\
&\leq \mathbb{P}_{\Omega}^{\infty}\Big(nm\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]})\big] > 0\Big).
\end{aligned}$$

Let $\Gamma \triangleq \{\mu \in \mathcal{P}_{\mathcal{X}} | f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \mu) > f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu)\}$. We have that

$$\begin{aligned}
&\mathbb{P}_{\Omega}^{\infty}\Big(nm\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,m]})\big] > 0\Big) \\
&= \mathbb{P}_{\Omega}^{\infty}\big\{\Pi_{\mathbf{X}^n[1,m]} \in \Gamma\big\} \\
&\leq \sum_{\substack{(U_1,\dots,U_K)\in\mathcal{P}_{mn_1}\times\dots\times\mathcal{P}_{mn_K} \\ \boldsymbol{\alpha}^T\boldsymbol{U}\in\Gamma}} e^{-\sum_{k=1}^K mn_k D(U_k||p_{0,k})} \\
&\leq \Big(\prod_k |\mathcal{P}_{mn_k}|\Big) \\
&\quad\cdot \exp\Bigg(-\inf_{\substack{(U_1,\dots,U_K)\in\mathcal{P}_{mn_1}\times\dots\times\mathcal{P}_{mn_K} \\ \boldsymbol{\alpha}^T\boldsymbol{U}\in\Gamma}} \sum_{k=1}^K mn_k D(U_k||p_{0,k})\Bigg) \\
&\leq \Big(\prod_k |\mathcal{P}_{mn_k}|\Big)e^{-hm},
\end{aligned} \tag{10}$$

where the last step is due to the fact that $\mathcal{P}_{mn_k} \in \mathcal{P}_{\mathcal{X}}$, $\forall k$. Note that $f_{\boldsymbol{P}}(\boldsymbol{\alpha}, \boldsymbol{Q}) \geq 0$ for any Q and the equality holds if and only if $\boldsymbol{\alpha}^T\boldsymbol{P} = \boldsymbol{Q}$ almost everywhere. We then have that $\boldsymbol{\alpha}^T\boldsymbol{P}_0 \notin \Gamma$ and $h > 0$. Therefore, for any Ω and $m \geq 1$,

$$\mathbb{P}_{\Omega}^{\infty}(Y > m) \leq \Big(\prod_k |\mathcal{P}_{mn_k}|\Big)e^{-mh}. \tag{11}$$

Define regeneration times $Y_0 = 0$ and for $r \geq 0$, $Y_{r+1} = \inf\big\{t > Y_r : W[t] \leq 0\big\}$. Let $R = \inf\{r : Y_r \leq \infty$ and $W[t] \geq b$ for some $Y_r < t \leq Y_{r+1}\}$ denote the index of the first cycle in which $W[t]$ crosses b. Note that according to the recursive update rule of $\hat{\nu}_t$ and $W[t]$, the test statistics in cycle $r+1$ are independent of the samples in cycles $1, \cdots, r$. For any Ω, we have that $\mathbb{E}_{\Omega}^{\infty}[\tau_e(b)] \geq \mathbb{E}_{\Omega}^{\infty}[R] = \sum_{r=0}^{\infty} \mathbb{P}_{\Omega}^{\infty}(R \geq r)$. For any Ω and $m \geq 1$, we have that

$$\begin{aligned}
&\mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < Y) \\
&= \mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < Y, Y \leq m) + \mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < Y, Y > m) \\
&\leq \mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < m) + \mathbb{P}_{\Omega}^{\infty}(Y > m).
\end{aligned} \tag{12}$$

Consider the first term in (12) $\mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < m)$:

$$\begin{aligned}
&\mathbb{P}_{\Omega}^{\infty}\big(\tau_e(b) < m\big) = \mathbb{P}_{\Omega}^{\infty}\Big(\max_{1\leq t<m} W[t] \geq b\Big) \\
&\leq \sum_{1\leq t<m} \mathbb{P}_{\Omega}^{\infty}\Big(W[t] \geq b\Big) \\
&= \sum_{1\leq t<m} \mathbb{P}_{\Omega}^{\infty}\Big(n\hat{t}\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]})\big] \geq b\Big).
\end{aligned}$$

Define $\Gamma_{b,t} \triangleq \big\{\mu \in \mathcal{P}_{\mathcal{X}} \big| n\hat{t}\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \mu) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu)\big] \geq b\big\}$. For all $\mu \in \Gamma_{b,t}$, we have that $n\hat{t}f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \mu) \geq b + n\hat{t}f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu) \geq b$, where the last inequality is due to the facts that $\hat{t} \geq 0$ and $f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \mu) \geq 0$. For any Ω and $1 \leq t < m$, following the same idea as (10), we have that

$$\begin{aligned}
&\mathbb{P}_{\Omega}^{\infty}\Big(n\hat{t}\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[\hat{\nu}_t,t]})\big] > b\Big) \\
&= \mathbb{P}_{\Omega}^{\infty}\Big\{\Pi_{\mathbf{X}^n[\hat{\nu}_t,t]} \in \Gamma_{b,t}\Big\} \\
&\leq \Big(\prod_k |\mathcal{P}_{\hat{t}n_k}|\Big) \exp\Bigg(-\inf_{\substack{(U_1,\dots,U_K)\in(\mathcal{P}_{\mathcal{X}})^K \\ \boldsymbol{\alpha}^T\boldsymbol{U}\in\Gamma_{b,t}}} \sum_{k=1}^K n_k\hat{t}D(U_k||p_{0,k})\Bigg) \\
&\leq \Big(\prod_k |\mathcal{P}_{mn_k}|\Big)e^{-b},
\end{aligned} \tag{13}$$

We then have that for any Ω, $\mathbb{P}_{\Omega}^{\infty}\big(\tau_e(b) < m\big) \leq m\left(\prod_k |\mathcal{P}_{mn_k}|\right)e^{-b}$. Let $m = \frac{b}{h}$, combing (11) and (13),

$$\mathbb{P}_{\Omega}^{\infty}(\tau_e(b) < Y) = \Big(\frac{b}{h}+1\Big)\Big(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\Big)e^{-b}. \tag{14}$$

It then follows that

$$\begin{aligned}
&\mathbb{P}_{\Omega}^{\infty}(R \geq r) = \mathbb{P}_{\Omega}^{\infty}\Big(W[t] < b, \forall Y_{m-1} \leq t \leq Y_m, \forall 1 \leq m \leq r\Big) \\
&= \prod_{m=1}^{r} \mathbb{P}_{\Omega}^{\infty}\Big(W[t] < b, \forall Y_{m-1} \leq t \leq Y_m\Big) \\
&\geq \Bigg(1 - \Big(\frac{b}{h}+1\Big)\Big(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\Big)e^{-b}\Bigg)^r,
\end{aligned} \tag{15}$$

where the second equality is due to (14) and the independence among the cycles [21]. Therefore, for any Ω,

$$\begin{aligned}
\mathbb{E}_{\Omega}^{\infty}[\tau_e(b)] &\geq \sum_{r=0}^{\infty} \Bigg(1 - \Big(\frac{b}{h}+1\Big)\Big(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\Big)e^{-b}\Bigg)^r \\
&= \frac{e^b}{\Big(\frac{b}{h}+1\Big)\Big(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\Big)}.
\end{aligned} \tag{16}$$

This completes the proof. □

To guarantee that $\inf_\Omega \mathbb{E}_\Omega^\infty\big[\tau_e(b)\big] \ge \gamma$, it suffices to choose b such that $\frac{e^b}{\left(\frac{b}{h}+1\right)\left(\prod_k |\mathcal{P}_{\frac{b}{h}n_k}|\right)} = \gamma$ and $b \sim \log \gamma$.

Note that an upper bound on the WADD for τ_e is difficult to obtain. To understand the detection delay of the proposed computationally efficient test, we then study the case when the change occurs at $\nu = 1$. We have the following result.

Proposition 1. *Consider the case with* $\nu = 1$. *Then, as* $t \to \infty$, $n\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]})\big] \to n f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1)$, *almost surely.*

Proof. According to the Glivenko–Cantelli theorem [20], as $t \to \infty$, under the post-change distribution, the empirical distribution $\Pi_{\mathbf{X}^n[1,t]}$ convergences to $\boldsymbol{\alpha}^T \boldsymbol{P}_1$ almost surely. Due to the fact that $f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1) = 0$, we have that $\frac{1}{t}(t-1+1)n\big[f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]}) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, \Pi_{\mathbf{X}^n[1,t]})\big]$ converges to $n f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1)$ almost surely. □

Intuitively, Proposition 1 implies that if the change is at $\nu = 1$ and regeneration does not happen, then the detection delay of the computationally efficient algorithm increases linearly with the threshold b at the rate of $1/(n f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1))$.

Let $\widetilde{\mathbb{P}}_0 = \frac{1}{|\mathcal{S}_{n,\lambda}|}\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{0,\sigma}$ and $\widetilde{\mathbb{P}}_1 = \frac{1}{|\mathcal{S}_{n,\lambda}|}\sum_{\sigma\in\mathcal{S}_{n,\lambda}} \mathbb{P}_{1,\sigma}$. We then present the following universal lower bound on the WADD, and show that the slope is also $1/(n f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1))$ when n is large.

Proposition 2. *For large* γ, *we have that* $\inf_{\tau:\text{WARL}\ge\gamma} \text{WADD}(\tau) \sim \frac{\log\gamma}{D(\widetilde{\mathbb{P}}_1\|\widetilde{\mathbb{P}}_0)}(1+o(1))$. *Moreover, as* $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{n} D\Big(\widetilde{\mathbb{P}}_1\Big|\Big|\widetilde{\mathbb{P}}_0\Big) = f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1). \tag{17}$$

Proof. It was shown in [1] that the mixture CuSum τ^* is exactly optimal for the QCD problem in Section II. Then, as $\gamma \to \infty$, we have that $\inf_{\tau:\text{WARL}\ge\gamma} \text{WADD}(\tau) = \text{WADD}(\tau^*)$. From Theorem 4 in [22] and the optimality of τ^* in [1], as $\gamma \to \infty$, it follows that $\text{WADD}(\tau^*) \sim \frac{\log\gamma}{D\left(\widetilde{\mathbb{P}}_1\middle|\middle|\widetilde{\mathbb{P}}_0\right)}(1+o(1))$. From Lemma 4, we have that for any $\sigma \in \mathcal{S}_{n,\lambda}$,

$$\log\frac{\widetilde{\mathbb{P}}_1(X^n)}{\widetilde{\mathbb{P}}_0(X^n)} = \log\frac{\mathbb{P}_{1,\sigma}\big(T(\Pi_{X^n})\big)}{\mathbb{P}_{0,\sigma}\big(T(\Pi_{X^n})\big)}. \tag{18}$$

Let $\mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_\theta, \epsilon) = \big\{\mu \in \mathcal{P}_\mathcal{X} : \sup_{x\in\mathcal{X}} \big|\mu(x) - \boldsymbol{\alpha}^T \boldsymbol{P}_\theta(x)\big| \le \epsilon\big\}$ denote the ball centered at $\boldsymbol{\alpha}^T \boldsymbol{P}_\theta$ with radius $\epsilon > 0$. According to the Glivenko–Cantelli theorem [20], we then have that for any $\sigma \in \mathcal{S}_{n,\lambda}$ and $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\theta,\sigma}\Big\{\sup_{x\in\mathcal{X}} \big|\Pi_{X^n}(x) - \boldsymbol{\alpha}^T \boldsymbol{P}_\theta(x)\big| > \epsilon\Big\} = 0. \tag{19}$$

It then follows that for any $\sigma \in \mathcal{S}_{n,\lambda}$ and $\epsilon > 0$,

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{P}_{\theta,\sigma}\Big\{\Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_\theta, \epsilon)\Big\}\\ &= \lim_{n\to\infty} \mathbb{P}_{\theta,\sigma}\Big\{\sup_{x\in\mathcal{X}} \big|\Pi_{X^n}(x) - \boldsymbol{\alpha}^T \boldsymbol{P}_\theta(x)\big| > \epsilon\Big\} = 0.\end{aligned} \tag{20}$$

It was shown in Lemma 5.3 in [15] that $f_{\boldsymbol{P}_\theta}(\boldsymbol{\alpha}, P)$ is a continuous function of P for any $\theta \in \{0, 1\}$. Therefore, $f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, P) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, P)$ is a continuous function of P. Then we have that for any $\epsilon > 0$, there exists an $\eta(\epsilon) > 0$ such that $\forall P \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)$,

$$\begin{aligned}&f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1) - \eta(\epsilon) < f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, P) - f_{\boldsymbol{P}_1}(\boldsymbol{\alpha}, P)\\ &< f_{\boldsymbol{P}_0}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^T \boldsymbol{P}_1) + \eta(\epsilon),\end{aligned} \tag{21}$$

where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$. We then have that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} D\Big(\widetilde{\mathbb{P}}_1\Big|\Big|\widetilde{\mathbb{P}}_0\Big)\\
&= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{\widetilde{\mathbb{P}}_1}\Big[\log \mathbb{P}_{1,\sigma}\big(T(\Pi_{X^n})\big) - \log \mathbb{P}_{0,\sigma}\big(T(\Pi_{X^n})\big)\Big]\\
&\overset{(a)}{\le} \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{\widetilde{\mathbb{P}}_1}\Bigg[\log\Big(\prod_k |\mathcal{P}_{n_k}|\Big) - \log\Big(\prod_{k=1}^K \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\\
&\quad - \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{1,k})\\
&\quad + \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{0,k})\Bigg]\\
&= \lim_{n\to\infty} \frac{1}{n}\widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\big)\mathbb{E}_{\widetilde{\mathbb{P}}_1}\Bigg[\log\Big(\prod_k |\mathcal{P}_{n_k}|\Big)\\
&\quad - \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{1,k})\\
&\quad + \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{0,k})\\
&\quad - \log\Big(\prod_{k=1}^K \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\Bigg|\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\Bigg]\\
&\quad + \lim_{n\to\infty} \frac{1}{n}\widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\big)\mathbb{E}_{\widetilde{\mathbb{P}}_1}\Bigg[\log\Big(\prod_k |\mathcal{P}_{n_k}|\Big)\\
&\quad - \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{1,k})\\
&\quad + \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{0,k})\\
&\quad - \log\Big(\prod_{k=1}^K \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\Bigg|\Pi_{X^n} \notin \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\Bigg]\\
&\overset{(b)}{=} \lim_{n\to\infty} \frac{1}{n}\widetilde{\mathbb{P}}_1\big(\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\big)\mathbb{E}_{\widetilde{\mathbb{P}}_1}\Bigg[\\
&\quad - \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{1,k})\\
&\quad + \inf_{\substack{(U_1,\ldots,U_K)\in\mathcal{P}_{n_1}\times\ldots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}} \sum_{k=1}^K n_k D(U_k||p_{0,k})
\end{aligned}$$

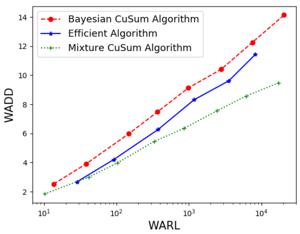

Fig. 1. Comparison of the three algorithms: $n = 2$.

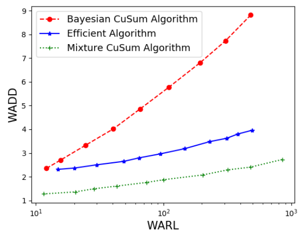

Fig. 2. Comparison of the three algorithms: $n = 8$.

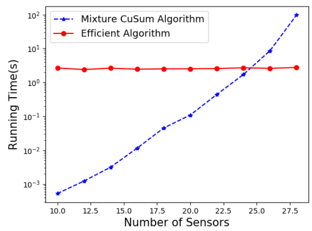

Fig. 3. Comparison of the computational complexity.

$$\begin{aligned}&\Big|\Pi_{X^n} \in \mathcal{B}(\boldsymbol{\alpha}^T \boldsymbol{P}_1, \epsilon)\Big]\\ &+ \lim_{n\to\infty} \frac{1}{n}\widetilde{\mathbb{P}}_1\left(\Pi_{X^n}\in\mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon)\right)\Bigg(\log\Big(\prod_k|\mathcal{P}_{n_k}|\Big)\\ &- \log\Big(\prod_{k=1}^K \frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\Bigg)\\ &\overset{(c)}{\le} f_{\boldsymbol{P}_0}(\boldsymbol{\alpha},\boldsymbol{\alpha}^T\boldsymbol{P}_1)+\eta(\epsilon),\end{aligned} \tag{22}$$

where the inequality (a) is due to the bound of the probability of type classes [17]: $\frac{1}{(n_k+1)^{|\mathcal{X}|}}2^{-n_k D(U_k||p_{\theta,k})} \le p_{\theta;k}^{\bigotimes n_k}(T_{n_k}(U_k)) \le 2^{-n_k D(U_k||p_{\theta,k})}$, the equality in (b) is due to the fact that $\lim_{n\to\infty}\widetilde{\mathbb{P}}_1(\Pi_{X^n}\notin\mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon)) = 0$ and the inequality (c) is due to (21) and the fact that $\lim_{n\to\infty}\frac{1}{n}\widetilde{\mathbb{P}}_1(\Pi_{X^n}\in\mathcal{B}(\boldsymbol{\alpha}^T\boldsymbol{P}_1,\epsilon))\Big(\log\Big(\prod_k|\mathcal{P}_{n_k}|\Big) - \log\Big(\prod_{k=1}^K\frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big)\Big) = 0.$

For the lower bound, following the same idea as in (22), we have that

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}D(\widetilde{\mathbb{P}}_1||\widetilde{\mathbb{P}}_0)\\ &\ge \lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\widetilde{\mathbb{P}}_1}\Bigg[\log\Big(\prod_{k=1}^K\frac{1}{(n_k+1)^{|\mathcal{X}|}}\Big) - \log\Big(\prod_k|\mathcal{P}_{n_k}|\Big)\\ &\quad - \inf_{\substack{(U_1,\dots,U_K)\in\mathcal{P}_{n_1}\times\dots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}}\sum_{k=1}^K n_k D(U_k||p_{1,k})\\ &\quad + \inf_{\substack{(U_1,\dots,U_K)\in\mathcal{P}_{n_1}\times\dots\times\mathcal{P}_{n_K}\\ \boldsymbol{\alpha}^T\boldsymbol{U}=\Pi_{X^n}}}\sum_{k=1}^K n_k D(U_k||p_{0,k})\Bigg]\\ &\ge f_{\boldsymbol{P}_0}(\boldsymbol{\alpha},\boldsymbol{\alpha}^T\boldsymbol{P}_1)-\eta(\epsilon).\end{aligned} \tag{23}$$

By (22) and (23), we have that $\lim_{n\to\infty}\frac{1}{n}D\Big(\widetilde{\mathbb{P}}_1||\widetilde{\mathbb{P}}_0\Big) = f_{\boldsymbol{P}_0}(\boldsymbol{\alpha},\boldsymbol{\alpha}^T\boldsymbol{P}_1)$. □

Combining Propositions 1 and 2, it can be seen that the tradeoff between the WADD and WARL for our computationally efficient test is close to the optimal one for large n. This demonstrates the advantage of our test that for large networks, it has a similar statistical efficiency comparing to the optimal test, and has a significantly reduced computational complexity.

V. Simulation Results

In this section, we provide some numerical results. We first consider a simple example with $n = 2$, $K = 2$, $n_1 = 1$ and $n_2 = 1$. The pre- and post-change distributions for group 1 are binomial distribution $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.3)$, respectively, and for group 2 are $\mathcal{B}(10, 0.5)$ and $\mathcal{B}(10, 0.7)$, respectively. We compare the performance of our efficient algorithm with the optimal mixture CuSum algorithm in [1] and a heuristic Bayesian CuSum algorithm $T_B = \inf\Big\{t \ge 1 : \max_{1\le j\le t}\sum_{i=j}^t \log\frac{\prod_{i=1}^n(\sum_{k=1}^K\frac{n_k}{n}p_{1,k}(X_i[t]))}{\prod_{i=1}^n(\sum_{k=1}^K\frac{n_k}{n}p_{0,k}(X_i[t]))} \ge b\Big\}$. The test statistics of these three algorithms are all symmetric, and therefore for different Ω's, the ADD and ARL are the same.

We then repeat the experiment for $n = 8$, $n_1 = 4$ and $n_2 = 4$ with the same distributions. For the two cases with $n = 2$ and $n = 8$, we plot the WADD as a function of the WARL in Fig. 1 and Fig. 2, respectively. It can be seen that mixture CuSum outperforms the other two tests, and our efficient test has a better performance than the intuitive Bayesian CuSum. More importantly, as n increases, the slope of the WADD-WARL tradeoff curve of the efficient algorithm is closer to the one of the optimal mixture CuSum algorithm. This conforms to the design of our efficient test which is to approximate the optimal mixture CuSum when n is large.

In Fig. 3, we show the computational efficiency of our algorithm. We compare the running time of one step update of our efficient algorithm and the optimal mixture CuSum algorithm. From Fig. 3, one can see that as n increases, the running time of the mixture CuSum increases exponentially, while the running time of our efficient test stays the same.

VI. Conclusion

In this paper, we studied the quickest change detection problem in anonymous heterogeneous sensor networks. We proposed a computationally efficient test to approximate the optimal mixture CuSum in [1]. We further developed its WARL lower bound for practical false alarm control. It remains to see whether our computationally efficient algorithm is also asymptotically optimal as $n \to \infty$.

References

[1] Z. Sun, S. Zou, and Q. Li, "Quickest change detection in anonymous heterogeneous sensor networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 5925–5929.

[2] J. Horey, M. M. Groat, S. Forrest, and F. Esponda, "Anonymous data collection in sensor networks," in *Proc. Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous)*. IEEE, 2007, pp. 1–8.

[3] M. M. Groat, W. Hey, and S. Forrest, "KIPDA: k-indistinguishable privacy-preserving data aggregation in wireless sensor networks," in *Proc. IEEE INFOCOM*, 2011, pp. 2024–2032.

[4] B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM SIGKDD Explorations Newsletter*, vol. 10, no. 2, pp. 12–22, 2008.

[5] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems," *Applied Sequential Methodologies: Real-World Examples with Data Analysis*, vol. 173, pp. 339–370, 2004.

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

[7] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.

[8] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *The Annals of Statistics*, pp. 670–692, 2013.

[9] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 3662–3675, 2016.

[10] V. Raghavan and V. V. Veeravalli, "Quickest change detection of a Markov process across a sensor array," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1961–1981, 2010.

[11] O. Hadjiliadis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3346–3359, 2009.

[12] M. Ludkovski, "Bayesian quickest detection in sensor arrays," *Sequential Analysis*, vol. 31, no. 4, pp. 481–504, 2012.

[13] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *IEEE Trans. Inform. Theory*, vol. 66, no. 4, pp. 2280–2295, 2020.

[14] L. Xie, S. Zou, Y. Xie, and V. V. Veeravalli, "Sequential (quickest) change detection: Classical results and new directions," *IEEE Journal on Selected Areas in Information Theory*, pp. 1–1, 2021.

[15] W. N. Chen and I. H. Wang, "Anonymous heterogeneous distributed detection: Optimal decision rules, error exponents, and the price of anonymity," *IEEE Trans. Inform. Theory*, 2019.

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

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 2006.

[18] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[19] Michael Grant and Stephen Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, pp. 95–110. Springer-Verlag, 2008.

[20] H. G. Tucker, "A generalization of the Glivenko-Cantelli theorem," *Ann. Math. Statist.*, vol. 30, no. 3, pp. 828–830, 1959.

[21] S. Asmussen, *Applied probability and queues*, vol. 51, Springer Science & Business Media, 2008.

[22] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.