2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517771

Quickest Dynamic Anomaly Detection in Anonymous Heterogeneous Sensor Networks

Zhongchang Sun and Shaofeng Zou
Electrical Enginnering
University at Buffalo, the State University of New York
Buffalo, NY, USA
Email: zhongcha@buffalo.edu, szou3@buffalo.edu

***Abstract*—The problem of quickest dynamic anomaly detection in anonymous heterogeneous sensor networks is studied. The n heterogeneous sensors can be divided into K types with different data generating distributions. At some unknown time, an anomaly emerges in the network and changes the data generating distribution of the sensors. The goal is to detect the anomaly as quickly as possible, subject to false alarm constraints. The anonymous setting is studied, where the fusion center does not know which sensor that each sample comes from, and thus does not know its exact distribution. Firstly, the static setting is investigated where the sensor affected by the anomaly does not change with time. A generalized mixture CuSum algorithm is constructed and is further shown to be asymptotically optimal. The problem is then extended to a dynamic setting where the sensor affected by the anomaly changes with time. An asymptotically optimal weighted mixture CuSum algorithm is proposed. Numerical results are also provided to validate the theoretical results.**

I. Introduction

In quickest change detection problems (QCD) [1]–[3], a decision maker collects samples sequentially. At some unknown time, a change occurs in the data generating distribution. The goal is to detect the change as quickly as possible subject to false alarm constraints. The QCD problem in sensor networks has been widely studied in the literature, e.g., [4]–[13]. In these papers, the sensor that each sample comes from is known. One CuSum algorithm can be implemented for each sensor, and be further combined to design algorithms with certain optimality guarantee. However, in many practical applications, sensors may be anonymous [14], i.e., the fusion center does not know which sensor that each sample comes from (see e.g., [15], [16] for anonymous data collection approaches). For example, in social settings [17], where human participants are involved, privacy and anonymity are required to protect the participants. In this paper, we investigate anonymous sensor networks. We consider a general setting with K types of heterogeneous sensors, and different types of sensors follow different data generating distributions.

In this paper, we investigate the problem of quickest dynamic anomaly detection in anonymous heterogeneous networks, where one unknown sensor is affected and undergoes a change in its data generating distribution after an anomaly emerges at some unknown time. The goal is to detect the anomaly as quickly as possible subject to false alarm constraints. We first investigate the static setting, where the sensor affected by the anomaly doesn't change with time, and then extend to the dynamic setting, where the sensor affected by the anomaly changes with time. For static anomaly, we construct a generalized mixture CuSum algorithm and prove that it is second-order asymptotically optimal. For dynamic anomaly, we construct a weighted mixture CuSum algorithm, and prove that it is first-order asymptotically optimal.

In [14], the binary hypothesis testing problem with anonymous sensors was studied and an optimal mixture likelihood ratio test (MLRT) was developed. In [18], the QCD problem with anonymous sensors was investigated, where all the sensors undergo a change in data generating distribution after the change. However, an anomaly may not affect all nodes, especially for large networks and in the distributed setting. In this paper, we focus on the case where the anomaly only affects one node. An anomaly may also be dynamic, and affect different nodes at different times, e.g., a moving target in surveillance system, spreading fake news in social network. Existing studies of quickly detecting dynamic change mostly focus on the non-anonymous setting, e.g., [12], [19], [20]. Our setting is similar to the one in [20] but in an anonymous setting. Our major technical challenge is due to the increased ambiguity of post-change distribution caused by anonymity.

Due to space limitation, we provide only part of the proofs, and the full proof can be found in [21].

II. Problem Formulation

Consider a network consisting of n heterogeneous sensors that can be divided into K types. Each type k has n_k sensors, $1 \leq k \leq K$. The distributions of the observations from type k sensors are $p_{\theta,k}$, $\theta \in \{0,1\}$. At some unknown time ν, an anomaly emerges and changes the data generating distributions of the affected sensor. If a node of type k is affected by the anomaly, then its samples are generated by $p_{1,k}$, otherwise, by $p_{0,k}$. The goal is to detect the anomaly as quickly as possible subject to false alarm constraints. The centralized setting with a fusion center is considered. The sensors are anonymous, i.e., the fusion center doesn't know the type of sensors that each observation comes from.

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

Denote by $X^n[t] = \{X_1[t], \dots, X_n[t]\}$ the n collected samples at time $t \geq 1$, which are assumed to be independent. We further assume that $X^n[t_1]$ is independent from $X^n[t_2]$ for any $t_1 \neq t_2$. Let $\mathbf{X}^n[t_1, t_2] = \{X^n[t_1], \cdots, X^n[t_2]\}$ for any $t_1 \leq t_2$. Let $\mathcal{K} = \{1, 2, \cdots, K\}$.

In this paper, we consider a dynamic anomaly, i.e., the sensor affected by the anomaly changes with time. After an anomaly emerges, one sensor of an unknown type is affected. Denote by $S[t] \in \mathcal{K} \cup \{0\}$ the type of the affected sensor at time t. We set $S[t] = 0$ for $t < \nu$, i.e., when there is no anomaly in the network. Let $\boldsymbol{S} \triangleq \{S[t]\}_{t=1}^{\infty}$ be the trajectory of the anomaly. Here $\boldsymbol{S}$ is *unknown* to the decision maker.

In order to distinguish among nodes being affected or not and also among nodes of different types, we construct $2K$ groups, and each group is associated with a distinct data generating distribution. Specifically, nodes in the first K groups are not affected by the anomaly, and nodes in the remaining K groups are affected by the anomaly. For nodes in group $k \in \mathcal{K}$, samples are generated by $p_{0,k}$, and for nodes in group $K < k \leq 2K$, their samples are generated by $p_{1,k-K}$.

Even if the unknown trajectory of the anomaly is given, the distribution of $X^n[t]$ can still not be fully specified due to the anonymity. To characterize the distribution of $X^n[t]$, we define a labeling function $\sigma_t^{S[t]} : \{1, \dots, n\} \to \{1, \dots, K, S[t] + K\}$ that maps each sample $X_i[t]$, $1 \leq i \leq n$, to a group $k \in \{1, 2, \dots, K, K + S[t]\}$. Specifically, if $\sigma_t^{S[t]}(i) = k$, then $X_i[t] \sim p_{0,k}$, for $k \in \mathcal{K}$, and $X_i[t] \sim p_{1,k-K}$, for $K < k \leq 2K$. We note that $\sigma_t^{S[t]}$ is *unknown* to the decision maker, and changes with time due to the anonymity.

Before the anomaly emerges $t < \nu$, there are n_k sensors in group k, $\forall k \in \mathcal{K}$, and 0 sensors in group k, $\forall K < k \leq 2K$. Then, there are in total $\binom{n}{n_1, \dots, n_K}$ possible $\sigma_t^{S[t]}$'s satisfying $|\{i : \sigma_t^{S[t]}(i) = k\}| = n_k \ \forall k \in \mathcal{K}$. We denote the collection of all such labelings by $\mathcal{S}_{n,\lambda_0}$, where $\lambda_0 = \{n_1, \dots, n_K\}$. After the anomaly emerges, i.e., $t \geq \nu$, one sensor of type $S[t] \neq 0$ is affected. Then, the number of sensors in group $S[t]$ and $S[t] + K$ are $n_{S[t]} - 1$ and 1, respectively. Then, there are $\binom{n}{n_1, \dots, n_{S[t]}-1, \dots, n_K, 1}$ possible $\sigma_t^{S[t]}$'s satisfying

$$|\{i : \sigma_t^{S[t]}(i) = k\}| = \begin{cases} n_k, & \text{if } k \in \mathcal{K} \backslash \{S[t]\}, \\ n_k - 1, & \text{if } k = S[t], \\ 1, & \text{if } k = S[t] + K, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

We then denote the collection of all such labelings by $\mathcal{S}_{n,\lambda_{S[t]}}$, where $\lambda_{S[t]} = \{n_1, \dots, n_{S[t]} - 1, \dots, n_K, 1\}$.

Before the anomaly emerges, i.e., $t < \nu$, we have

$$\mathbb{P}_{0,\sigma_t^0}(X^n[t]) = \prod_{i=1}^{n} p_{0,\sigma_t^0(i)}(X_i[t]), \quad (2)$$

for some unknown $\sigma_t^0 \in \mathcal{S}_{n,\lambda_0}$. At time $t \geq \nu$, we have

$$\begin{aligned} \mathbb{P}_{S[t],\sigma_t^{S[t]}}(X^n[t]) = & \prod_{i:\sigma_t^{S[t]}(i) \leq K} p_{0,\sigma_t^{S[t]}(i)}(X_i[t]) \\ & \times \prod_{i:\sigma_t^{S[t]}(i) > K} p_{1,\sigma_t^{S[t]}(i)-K}(X_i[t]), \quad (3) \end{aligned}$$

for some unknown $\sigma_t^{S[t]} \in \mathcal{S}_{n,\lambda_{S[t]}}$.

Let $\Omega_{\boldsymbol{S}} = \{\sigma_1^{S[1]}, ..., \sigma_\infty^{S[\infty]}\}$ be the collection of group labelings when the trajectory of the anomaly is $\boldsymbol{S}$. Let $\mathbb{P}_{\Omega_{\boldsymbol{S}}}^{\boldsymbol{S},\nu}$ denote the probability measure when the change point is at ν and the samples are generated according to (2), (3) and $\Omega_{\boldsymbol{S}}$, and let $\mathbb{E}_{\Omega_{\boldsymbol{S}}}^{\boldsymbol{S},\nu}$ denote the corresponding expectation. We then extend Lorden's criterion [22] and define the worst-case average detection delay (WADD) and the worst-case average running length (WARL) for any stopping time τ:

$$\begin{aligned} & \text{WADD}(\tau) \\ & = \sup_{\nu \geq 1} \sup_{\boldsymbol{S}} \sup_{\Omega_{\boldsymbol{S}}} \text{esssup} \mathbb{E}_{\Omega_{\boldsymbol{S}}}^{\boldsymbol{S},\nu} [(\tau - \nu)^+ | \mathbf{X}^n[1, \nu - 1]], \quad (4) \\ & \text{WARL}(\tau) = \inf_{\Omega} \mathbb{E}_{\Omega}^{\infty}[\tau], \quad (5) \end{aligned}$$

where $\Omega = \Omega_{\boldsymbol{S}}$ with $S[t] = 0, \forall t \geq 1$. The goal is to design a stopping rule that minimizes the WADD subject to a constraint on the WARL:

$$\inf_{\tau:\text{WARL}(\tau) \geq \gamma} \text{WADD}(\tau), \quad (6)$$

where $\gamma > 0$ is a pre-specified threshold.

III. A SIMPLE CASE: STATIC ANOMALY

We first consider a setting where the anomaly is static, i.e., the sensor affected by the anomaly does not change with time after the anomaly emerges. In this case, $S[t] = k$, $\forall t \geq \nu$. Let $\Omega_k = \{\sigma_1^0, \dots, \sigma_{\nu-1}^0, \sigma_\nu^k, \dots, \sigma_\infty^k\}$ be the corresponding group labellings. Then, $\text{WADD}(\tau) = \sup_{\nu \geq 1} \sup_k \sup_{\Omega_k} \text{esssup} \mathbb{E}_{\Omega_k}^{k,\nu}[(\tau - \nu)^+ | \mathbf{X}^n[1, \nu - 1]]$.

A. Universal Lower Bound on WADD

Let I_k denote the Kullback-Leibler (KL) divergence between two mixture distributions $\widetilde{\mathbb{P}}_k = \frac{1}{|\mathcal{S}_{n,\lambda_k}|} \sum_{\sigma^k \in \mathcal{S}_{n,\lambda_k}} \mathbb{P}_{k,\sigma^k}$ and $\widetilde{\mathbb{P}}_0 = \frac{1}{|\mathcal{S}_{n,\lambda_0}|} \sum_{\sigma^0 \in \mathcal{S}_{n,\lambda_0}} \mathbb{P}_{0,\sigma^0}$. Let $I^* = \min_{1 \leq k \leq K} I_k$. We then have the following theorem.

Theorem 1. *As $\gamma \to \infty$, we have that*

$$\inf_{\tau:WARL(\tau) \geq \gamma} WADD(\tau) \geq \frac{\log \gamma}{I^*} + O(1). \quad (7)$$

Proof. Consider a simple QCD problem with samples independent and identically distributed (i.i.d.) according to a pre-change distribution $\widetilde{\mathbb{P}}_0$ and a post-change distribution $\widetilde{\mathbb{P}}_k$, respectively. For this pair of pre- and post-change distributions, define the $\widetilde{\text{WADD}}_k$ and $\widetilde{\text{ARL}}$ for any stopping rule τ as follows:

$$\begin{aligned} \widetilde{\text{WADD}}_k(\tau) &= \sup_{\nu \geq 1} \text{ess sup} \widetilde{\mathbb{E}}_k^{\nu}[(\tau - \nu)^+ | \widetilde{\mathbf{X}}^n[1, \nu - 1]], \\ \widetilde{\text{ARL}}(\tau) &= \widetilde{\mathbb{E}}^{\infty}[\tau], \quad (8) \end{aligned}$$

where $\widetilde{\mathbb{E}}_k^{\nu}$ denotes the expectation when the change is at ν, the pre- and post-change distributions are $\widetilde{\mathbb{P}}_0$ and $\widetilde{\mathbb{P}}_k$, $\widetilde{\mathbf{X}}^n[t]$ for $1 \leq t \leq \nu - 1$ are i.i.d. from $\widetilde{\mathbb{P}}_0$, and $\widetilde{\mathbb{E}}^{\infty}$ denotes the expectation when there is no change and samples are generated according to $\widetilde{\mathbb{P}}_0$.

For any $k \in \mathcal{K}$, consider another QCD problem with samples distributed according to the pre-change distribution $\mathbb{P}_{0,\sigma_t^0}$ and the post-change distribution $\mathbb{P}_{k,\sigma_t^k}$, respectively. For this pair of pre- and post-change distributions, define the WADD_k and WARL for any stopping rule τ as follows:

$$\begin{aligned}\text{WADD}_k(\tau) &= \sup_{\nu\geq 1}\sup_{\Omega_k}\text{esssup}\mathbb{E}_{\Omega_k}^{k,\nu}[(\tau-\nu)^+|\mathbf{X}^n[1,\nu-1]],\\ \text{WARL}(\tau) &= \inf_{\Omega}\mathbb{E}_{\Omega}^{\infty}[\tau].\end{aligned}\tag{9}$$

For any $k \in \mathcal{K}$ and any τ satisfying $\text{WARL}(\tau) \geq \gamma$,

$$\begin{aligned}\text{WADD}(\tau) &\geq \sup_{\nu\geq 1}\sup_{\Omega_k}\text{esssup}\mathbb{E}_{\Omega_k}^{k,\nu}[(\tau-\nu)^+|\mathbf{X}^n[1,\nu-1]]\\ &\geq \sup_{\nu\geq 1}\text{ess}\sup\widetilde{\mathbb{E}}_k^{\nu}[(\tau-\nu)^+|\widetilde{\mathbf{X}}^n[1,\nu-1]]\\ &= \widetilde{\text{WADD}}_k(\tau),\end{aligned}\tag{10}$$

where the second inequality is due to the fact that for any τ, $\text{WADD}_k(\tau) \geq \widetilde{\text{WADD}}_k(\tau)$, which was proved in (18) of [18]. It was also shown in (18) of [18] that for any τ, $\text{WARL}(\tau) \leq \widetilde{\text{ARL}}(\tau)$. It then follows that

$$\begin{aligned}\inf_{\tau:\text{WARL}(\tau)\geq\gamma}\text{WADD}(\tau) &\geq \inf_{\tau:\widetilde{\text{ARL}}(\tau)\geq\gamma}\widetilde{\text{WADD}}_k(\tau)\\ &\geq \frac{\log\gamma}{I_k}+O(1), \text{ as } \gamma\to\infty,\end{aligned}\tag{11}$$

where the last inequality is due to the universal lower bound on WADD for a simple QCD problem [23]. □

B. Generalized Mixture CuSum

Since k is unknown, we use its maximum likelihood estimate. Let $W[t] = \max_{1\leq j\leq t}\max_{k\in\mathcal{K}}\sum_{i=j}^{t}\log\frac{\widetilde{\mathbb{P}}_k(X^n[i])}{\widetilde{\mathbb{P}}_0(X^n[i])}$. We then define the generalized mixture CuSum stopping time as: $T_G = \inf\{t : W[t] \geq b\}$, where $b > 0$ is the threshold. To compute $W[t]$ efficiently, we can keep K CuSums in parallel and take their maximum as $W[t]$.

Theorem 2. *Let* $b = \log(K\gamma)$, *then* $\textit{WARL}(T_G) \geq \gamma$. *As* $\gamma \to \infty$, $\textit{WADD}(T_G) \leq \frac{\log\gamma}{I_k^*} + O(1)$. T_G *is second-order asymptotically optimal.*

Proof for upper bound on WADD. Consider the mixture CuSum for problem in (9):

$$T_k = \inf\left\{t : \max_{1\leq j\leq t}\sum_{i=j}^{t}\log\frac{\widetilde{\mathbb{P}}_k(X^n[i])}{\widetilde{\mathbb{P}}_0(X^n[i])} \geq b\right\}.\tag{12}$$

It then follows that for any $1 \leq k \leq K$,

$$\begin{aligned}\text{WADD}_k(T_G) &= \sup_{\nu\geq 1}\sup_{\Omega_k}\text{esssup}\mathbb{E}_{\Omega_k}^{k,\nu}[(T_G-\nu)^+|\mathbf{X}^n[1,\nu-1]]\\ &\leq \sup_{\nu\geq 1}\sup_{\Omega_k}\text{esssup}\mathbb{E}_{\Omega_k}^{k,\nu}[(T_k-\nu)^+|\mathbf{X}^n[1,\nu-1]]\\ &\leq \frac{\log b}{I_k}+O(1),\end{aligned}\tag{13}$$

where the last equality is because of the exact optimality of the mixture CuSum algorithm (see Theorem 1 in [18]). Let $b = \log K\gamma$. We then have that

$$\begin{aligned}\text{WADD}(T_G) &= \sup_{k\in[1,K]}\text{WADD}_k(T_G)\\ &\leq \sup_{k\in[1,K]}\text{WADD}_k(T_k)\\ &= \sup_{k\in[1,K]}\frac{\log K\gamma}{I_k}+O(1)\\ &= \frac{\log\gamma}{I_k^*}+\frac{\log K}{I_k^*}+O(1), \text{ as } \gamma\to\infty.\end{aligned}\tag{14}$$

□

IV. Quickest Dynamic Anomaly Detection

A. Universal Lower Bound on WADD

Define the following weighted mixture distribution: $\widetilde{\mathbb{P}}_{\boldsymbol{\beta}}(X^n) = \sum_{k=1}^{K}\beta_k\widetilde{\mathbb{P}}_k(X^n)$, where $0 \leq \beta_k \leq 1$ and $\sum_{k=1}^{K}\beta_k = 1$. Denote by $I_{\boldsymbol{\beta}}$ the KL divergence between $\widetilde{\mathbb{P}}_{\boldsymbol{\beta}}$ and $\widetilde{\mathbb{P}}_0$. Let $\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} I_{\boldsymbol{\beta}}$.

For the universal lower bound on WADD, we have the following theorem.

Theorem 3. *As* $\gamma \to \infty$, *we have that*

$$\inf_{\tau:\textit{WARL}(\tau)\geq\gamma}\textit{WADD}(\tau) \geq \frac{\log\gamma}{I_{\boldsymbol{\beta}^*}}(1+o(1)).\tag{15}$$

B. Weighted Mixture CuSum

Define the log of weighted mixture likelihood ratio using $\boldsymbol{\beta}^*$: $\ell_{\boldsymbol{\beta}^*}(X^n) = \log\frac{\widetilde{\mathbb{P}}_{\boldsymbol{\beta}^*}(X^n)}{\widetilde{\mathbb{P}}_0(X^n)}$. The following property of $\boldsymbol{\beta}^*$ plays an important role in the proof of asymptotic optimality.

Lemma 1. *For any* $k \in \mathcal{K}$, $\mathbb{E}_{\widetilde{\mathbb{P}}_k}\left[\log\frac{\widetilde{\mathbb{P}}_{\boldsymbol{\beta}^*}(X^n)}{\widetilde{\mathbb{P}}_0(X^n)}\right] \geq I_{\boldsymbol{\beta}^*}$.

We then construct the following weighted mixture CuSum algorithm:

$$T_{\boldsymbol{\beta}^*}(b) = \inf\left\{t : \max_{1\leq j\leq t+1}\sum_{i=j}^{t}\ell_{\boldsymbol{\beta}^*}(X^n[i]) \geq b\right\},\tag{16}$$

for some positive threshold b.

The following theorem establishes the first-order asymptotic optimality of $T_{\boldsymbol{\beta}^*}$.

Theorem 4. *Let* $b = \log\gamma$, *then* $\textit{WARL}(T_{\boldsymbol{\beta}^*}) \geq b$. *Assume that* $\max_{k\in\mathcal{K}}\mathbb{E}_{\widetilde{\mathbb{P}}_k}\left[\ell_{\boldsymbol{\beta}^*}(X^n)^2\right] < \infty$. *As* $\gamma \to \infty$, $\textit{WADD}(T_{\boldsymbol{\beta}^*}) \leq \frac{\log\gamma}{I_{\boldsymbol{\beta}^*}}(1+o(1))$. *Thus,* $T_{\boldsymbol{\beta}^*}$ *is asymptotically optimal up to the first-order.*

Proof. For any permutation $\pi(X^n) = (X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$, we have that $\ell_{\boldsymbol{\beta}^*}(X^n) = \ell_{\boldsymbol{\beta}^*}(\pi(X^n))$. For any π, let $\hat{\sigma}^k = \sigma^k \circ \pi$. Then $\mathbb{E}_{k,\sigma^k}[\ell_{\boldsymbol{\beta}^*}(\pi(X^n))] = \mathbb{E}_{k,\sigma^k\circ\pi}[\ell_{\boldsymbol{\beta}^*}(X^n)] = \mathbb{E}_{k,\hat{\sigma}^k}[\ell_{\boldsymbol{\beta}^*}(X^n)]$. For any $\hat{\sigma}^k \in \mathcal{S}_{n,\lambda_k}$, a π can always be found so that $\sigma^k \circ \pi = \hat{\sigma}^k$. Thus, for any $\sigma^k, \hat{\sigma}^k \in \mathcal{S}_{n,\lambda_k}$, $\mathbb{E}_{k,\hat{\sigma}^k}[\ell_{\boldsymbol{\beta}^*}(X^n)] = \mathbb{E}_{k,\sigma^k}[\ell_{\boldsymbol{\beta}^*}(X^n)]$.

$T_{\boldsymbol{\beta}^*}$ is the same for any group labellings of $X^n[i]$ in $\mathcal{S}_{n,\lambda_k}$. Therefore, for any trajectory $\boldsymbol{S}$, we have that

$$\mathbb{E}^{\boldsymbol{S}}_{\sigma_1^{S[1]},\ldots,\sigma_i^{S[i]},\ldots,\sigma_\infty^{S[\infty]}}[T_{\boldsymbol{\beta}^*}] = \mathbb{E}^{\boldsymbol{S}}_{\sigma_1^{S[1]},\ldots,\hat{\sigma}_i^{S[i]},\ldots,\sigma_\infty^{S[\infty]}}[T_{\boldsymbol{\beta}^*}].$$

Due to the fact that the test statistic $\max_{1\le k\le t+1}\sum_{i=k}^{t}\ell_{\boldsymbol{\beta}^*}(X_i^n)$ has initial value 0 and remains non-negative, the delay is largest when the change happens at $\nu = 0$. Therefore, for any $\boldsymbol{S}$, we have that

$$\begin{aligned}&\text{WADD}_{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*})\\&=\sup_{\nu\ge 0}\sup_{\Omega_{\boldsymbol{S}}}\text{esssup}\mathbb{E}^{\boldsymbol{S},\nu}_{\Omega_{\boldsymbol{S}}}\left[(T_{\boldsymbol{\beta}^*}-\nu)^+ \mid \mathbf{X}^n[1,\nu-1]\right]\\&=\sup_{\Omega_{\boldsymbol{S}}}\mathbb{E}^{\boldsymbol{S},0}_{\Omega_{\boldsymbol{S}}}[T_{\boldsymbol{\beta}^*}].\end{aligned}\tag{17}$$

For any $T\ge \nu+1$, we have that

$$\begin{aligned}&\sup_{\substack{\{\sigma_1^{S[1]},\cdots,\sigma_T^{S[T]}\}\\ \in\mathcal{S}_{n,\lambda_{S[1]}}\times,\cdots,\times\mathcal{S}_{n,\lambda_{S[T]}}}}\sum_{t=1}^{T} t\mathbb{P}^{\boldsymbol{S},0}_{\sigma_1^{S[1]},\cdots,\sigma_T^{S[T]}}(T_{\boldsymbol{\beta}^*}=t)\\&=\sum_{t=1}^{T} t\frac{1}{\mid\mathcal{S}_{n,\lambda_{S[1]}}\mid\times\cdots\times\mid\mathcal{S}_{n,\lambda_{S[T]}}\mid}\\&\quad\times\sum_{\substack{\{\sigma_1^{S[1]},\cdots,\sigma_T^{S[T]}\}\\ \in\mathcal{S}_{n,\lambda_{S[1]}}\times,\cdots,\times\mathcal{S}_{n,\lambda_{S[T]}}}}\mathbb{P}^{\boldsymbol{S},0}_{\sigma_1^{S[1]},\cdots,\sigma_T^{S[T]}}(T_{\boldsymbol{\beta}^*}=t)\\&=\sum_{t=1}^{T} t\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*}=t).\end{aligned}\tag{18}$$

As $T\to\infty$, we have that

$$\sup_{\Omega_{\boldsymbol{S}}}\mathbb{E}^{\boldsymbol{S},0}_{\Omega_{\boldsymbol{S}}}[T_{\boldsymbol{\beta}^*}]=\widetilde{\mathbb{E}}_1^{\boldsymbol{S}}[T_{\boldsymbol{\beta}^*}]=\widetilde{\text{WADD}}_{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*}).\tag{19}$$

For any $\boldsymbol{S}$, we have $\text{WADD}_{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*})=\widetilde{\text{WADD}}_{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*})$. Therefore, $\text{WADD}(T_{\boldsymbol{\beta}^*})=\widetilde{\text{WADD}}(T_{\boldsymbol{\beta}^*})$ by taking sup over $\boldsymbol{S}$ on both sides. It then follows that

$$\text{WADD}(T_{\boldsymbol{\beta}^*})=\widetilde{\text{WADD}}(T_{\boldsymbol{\beta}^*})=\sup_{\boldsymbol{S}}\widetilde{\mathbb{E}}_1^{\boldsymbol{S}}[T_{\boldsymbol{\beta}^*}].\tag{20}$$

Let $0<\epsilon<I_{\boldsymbol{\beta}^*}$ and $n_b=\frac{b}{I_{\boldsymbol{\beta}^*}-\epsilon}$. For any trajectory $\boldsymbol{S}$, from the sum-integral inequality, we have that

$$\begin{aligned}\widetilde{\mathbb{E}}_1^{\boldsymbol{S}}\Big[\frac{T_{\boldsymbol{\beta}^*}}{n_b}\Big]&=\int_0^\infty\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\Big(\frac{T_{\boldsymbol{\beta}^*}}{n_b}>x\Big)\mathrm{d}x\\&\le\sum_{t=1}^{\infty}\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*}>tn_b)+1.\end{aligned}\tag{21}$$

For any $\boldsymbol{S}$, we have that

$$\begin{aligned}&\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*}>tn_b)=\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\max_{1\le k\le tn_b}\max_{1\le i\le k}\sum_{j=i}^{k}\ell_{\boldsymbol{\beta}^*}(X_j^n)<b\right)\\&\le\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\max_{1\le i\le mn_b}\sum_{j=i}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)<b,\forall m\in[t]\right)\\&\le\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)<b,\forall m\in[t]\right)\\&=\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}<I_{\boldsymbol{\beta}^*}-\epsilon,\forall m\in[t]\right)\\&=\prod_{m=1}^{t}\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}<I_{\boldsymbol{\beta}^*}-\epsilon\right).\end{aligned}\tag{22}$$

It then follows that

$$\begin{aligned}&\sup_{\boldsymbol{S}}\sum_{t=1}^{\infty}\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}(T_{\boldsymbol{\beta}^*}>tn_b)\\&\le\sup_{\boldsymbol{S}}\sum_{t=1}^{\infty}\prod_{m=1}^{t}\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}<I_{\boldsymbol{\beta}^*}-\epsilon\right).\end{aligned}\tag{23}$$

Then we will bound $\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}<I_{\boldsymbol{\beta}^*}-\epsilon\right)$.

Let $I_{\boldsymbol{S}_m}=\widetilde{\mathbb{E}}_1^{\boldsymbol{S}}\left[\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}\right]$. From 1, we have that

$$\begin{aligned}I_{\boldsymbol{S}_m}&=\widetilde{\mathbb{E}}_1^{\boldsymbol{S}}\left[\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}\right]\\&=\sum_{j=(m-1)n_b+1}^{mn_b}\mathbb{E}_{\widetilde{\mathbb{P}}_{S[j]}}\left[\frac{\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}\right]\\&=\frac{1}{n_b}\sum_{j=(m-1)n_b+1}^{mn_b}\mathbb{E}_{\widetilde{\mathbb{P}}_{S[j]}}\left[\ell_{\boldsymbol{\beta}^*}(X_j^n)\right]\ge I_{\boldsymbol{\beta}^*}.\end{aligned}\tag{24}$$

It then follows that for any $\boldsymbol{S}$ and m

$$\begin{aligned}&\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}<I_{\boldsymbol{\beta}^*}-\epsilon\right)\\&\le\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}<I_{\boldsymbol{S}_m}-\epsilon\right)\\&\le\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\left|\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}-I_{\boldsymbol{S}_m}\right|>\epsilon\right).\end{aligned}\tag{25}$$

Let $\sigma^2 = \max_{k\in\mathcal{K}}\text{Var}_{\widetilde{\mathbb{P}}_k}(\ell_{\boldsymbol{\beta}^*}(X^n))$. Since $\max_{k\in\mathcal{K}}\mathbb{E}_{\widetilde{\mathbb{P}}_k}\left[\ell_{\boldsymbol{\beta}^*}(X^n)^2\right]<\infty$, by Chebychev's inequality,

$$\begin{aligned}&\widetilde{\mathbb{P}}_1^{\boldsymbol{S}}\left(\left|\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}-I_{\boldsymbol{S}_m}\right|>\epsilon\right)\\&\le\text{Var}_{\widetilde{\mathbb{P}}_{\boldsymbol{S}}}\left(\frac{\sum_{j=(m-1)n_b+1}^{mn_b}\ell_{\boldsymbol{\beta}^*}(X_j^n)}{n_b}\right)\frac{1}{\epsilon^2}\\&=\frac{1}{\epsilon^2 n_b^2}\sum_{j=(m-1)n_b+1}^{mn_b}\text{Var}_{\widetilde{\mathbb{P}}_{S[j]}}(\ell_{\boldsymbol{\beta}^*}(X_j^n))\end{aligned}$$

$$\leq \frac{\sum_{j=(m-1)n_b+1}^{mn_b} \sigma^2}{n_b^2 \epsilon^2} = \frac{\sigma^2}{n_b \epsilon^2}. \quad (26)$$

Let $\delta = \frac{\sigma^2}{n_b \epsilon^2}$. From (21) and (26), we have that

$$\sup_{\boldsymbol{S}} \widetilde{\mathbb{E}}_1^{\boldsymbol{S}} \left[\frac{T_{\boldsymbol{\beta}^*}}{n_b} \right] \leq 1 + \sup_{\boldsymbol{S}} \sum_{t=1}^{\infty} \widetilde{\mathbb{P}}_1^{\boldsymbol{S}} (T_{\boldsymbol{\beta}^*} > t n_b)$$
$$\leq 1 + \sum_{t=1}^{\infty} (\frac{\sigma^2}{n_b \epsilon^2})^t = 1 + \sum_{t=1}^{\infty} \delta^t = \frac{1}{1-\delta}. \quad (27)$$

Therefore, we have

$$\sup_{\boldsymbol{S}} \widetilde{\mathbb{E}}_1^{\boldsymbol{S}} \left[T_{\boldsymbol{\beta}^*} \right] \leq \frac{b}{(I_{\boldsymbol{\beta}^*} - \epsilon)(1-\delta)}. \quad (28)$$

(28) holds for all ϵ. As $b \to \infty, \delta \to 0$. It then follows that

$$\text{WADD}(T_{\boldsymbol{\beta}^*}) = \sup_{\boldsymbol{S}} \widetilde{\mathbb{E}}_1^{\boldsymbol{S}} \left[T_{\boldsymbol{\beta}^*} \right] \leq \frac{b}{I_{\boldsymbol{\beta}^*}} (1 + o(1)). \quad (29)$$

We then show the ARL lower bound. For any $T \geq 1$, we have

$$\inf_{\substack{\{\sigma_1^0,\ldots,\sigma_T^0\} \\ \in \mathcal{S}_{n,\lambda_0}{}^{\otimes T}}} \sum_{t=1}^{T} t \mathbb{P}^{\infty}_{\sigma_1^0,\ldots,\sigma_T^0} (T_{\boldsymbol{\beta}^*} = t)$$
$$= \sum_{t=1}^{T} t \frac{1}{\mid \mathcal{S}_{n,\lambda_0} \mid^T} \sum_{\substack{\{\sigma_1^0,\ldots,\sigma_T^0\} \\ \in \mathcal{S}_{n,\lambda_0}{}^{\otimes T}}} \mathbb{P}^{\infty}_{\sigma_1^0,\ldots,\sigma_T^0} (T_{\boldsymbol{\beta}^*} = t)$$
$$= \sum_{t=1}^{T} t \widetilde{\mathbb{P}}^{\infty} (T_{\boldsymbol{\beta}^*} = t). \quad (30)$$

As $T \to \infty$, we have that

$$\text{WARL}(T_{\boldsymbol{\beta}^*}) = \widetilde{\text{ARL}}(T_{\boldsymbol{\beta}^*}). \quad (31)$$

$T_{\boldsymbol{\beta}^*}$ is the CuSum algorithm for a simple QCD problem with pre-change distribution $\widetilde{\mathbb{P}}_0$ and post-change distribution $\widetilde{\mathbb{P}}_{\boldsymbol{\beta}^*}$. From the optimal property of CuSum algorithm in [22] and [24], we have that when $b = \log \gamma$,

$$\text{WARL}(T_{\boldsymbol{\beta}^*}) = \widetilde{\text{ARL}}(T_{\boldsymbol{\beta}^*}) \geq \gamma. \quad (32)$$

Combining with Theorem 3, this shows that $T_{\boldsymbol{\beta}^*}$ is asymptotically optimal. □

V. Simulation Results

We set $n = 2$ and $K = 2$. For type I sensors, the pre- and post-change distributions are $\mathcal{B}(10, 0.3)$ and $\mathcal{B}(10, 0.4)$, respectively, where $\mathcal{B}$ denotes binomial distribution. For type II sensors, the pre- and post-change distributions are $\mathcal{B}(10, 0.8)$ and $\mathcal{B}(10, 0.6)$, respectively. Due to the difficulty of searching over all possible trajectories computationally, we plot the ADD and ARL for some randomly generated trajectories.

For the static setting, we compare our generalized mixture CuSum algorithm with a Bayesian mixture CuSum algorithm $T_B = \inf\{t : \max_{1 \leq j \leq t} \sum_{i=j}^{t} \log \frac{\frac{1}{2}\widetilde{\mathbb{P}}_1(X^n[i]) + \frac{1}{2}\widetilde{\mathbb{P}}_2(X^n[i])}{\widetilde{\mathbb{P}}_0(X^n[i])} \geq b\}$. One senor of type two is affected. In Fig. 1, we plot the ADD as function of ARL. It can be seen that our generalized mixture CuSum outperforms the Bayesian algorithm. Moreover, the relationship between the ADD and log of the ARL is linear.

For dynamic anomaly detection, we first compare our optimal weighted mixture CuSum algorithm with an arbitrarily weighted mixture CuSum, i.e., replace $\boldsymbol{\beta}^*$ in (16) with some arbitrarily $\boldsymbol{\beta}$, e.g., $\boldsymbol{\beta} = (\frac{1}{2}, \frac{1}{2})$. In Fig. 2, we plot the ADD as a function of ARL. It can be seen that our optimal weighted mixture CuSum algorithm outperforms the Bayesian weighted mixture CuSum algorithm.

We then compare the performance of our weighted mixture CuSum algorithm under two different trajectories. In Fig. 3, we plot the ADD as function of ARL. It can be seen that for two different trajectories, our optimal weighted mixture CuSum algorithm has approximately the same performance.

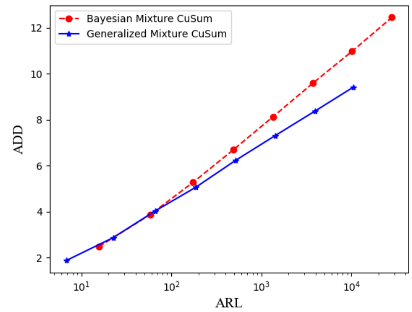

Fig. 1. Comparison of the Generalized Mixture CuSum Algorithm and A Bayesian Mixture CuSum Algorithm: Static Anomaly.

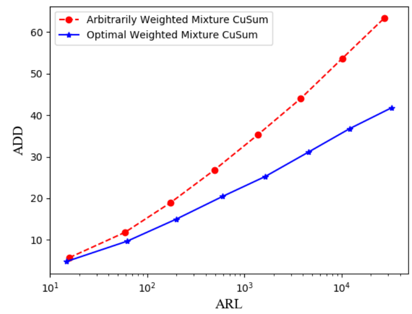

Fig. 2. Comparison of the Optimal Weighted CuSum Algorithm and An Arbitrarily Weighted One: Dynamic Anomaly.

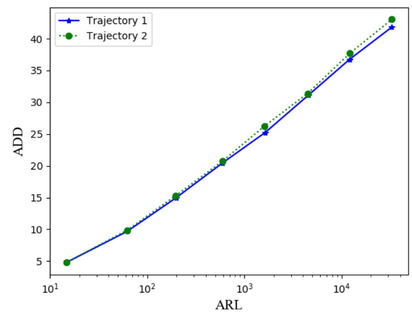

Fig. 3. Comparison of the Optimal Weighted CuSum under Two Different Trajectories.

REFERENCES

[1] V. V. Veeravalli and T. Banerjee, "Quickest change detection," *Academic press library in signal processing: Array and statistical signal processing*, vol. 3, pp. 209–256, 2013.

[2] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*, CRC Press, 2014.

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

[4] 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.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Sequential algorithms for moving anomaly detection in networks," *Sequential Analysis*, vol. 39, no. 1, pp. 6–31, 2020.

[20] G. Rovatsos, V. V. Veeravalli, and G. V. Moustakides, "Quickest detection of a dynamic anomaly in a heterogeneous sensor network," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 1171–1176.

[21] Z. Sun and S. Zou, "Quickest dynamic anomaly detection in anonymous heterogeneous sensor networks," available at https://drive.google.com/file/d/1WxHE4pZ5tk4pMKWVTWUG3PZNX1wXvZXc/view?usp=sharing, 2020.

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

[23] 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.

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