Quickest Detection of a Dynamic Anomaly in a Heterogeneous Sensor Network

Georgios Rovatsos[1], Venugopal V. Veeravalli[1], and George V. Moustakides[2]
[1]Department of Electrical and Computer Engineering, Coordinated Science Lab, University of Illinois at Urbana-Champaign
[2]Department of Electrical and Computer Engineering, University of Patras, Greece
Email: rovatso2@illinois.edu, vvv@illinois.edu, moustaki@upatras.gr

***Abstract*—The problem studied is one of quickest detection of an anomaly that emerges in a sensor network, and which may move across the network after it emerges. Each sensor in the network is characterized by a *non-anomalous* and an *anomalous* data-generating distribution, and these distributions could be different across the sensors. Initially, the observations at all the sensors are generated according to their corresponding non-anomalous distribution. After some *unknown* but *deterministic* time instant, a dynamic anomaly emerges in the network, affecting a different sensor as time progresses. The observations generated by the affected sensor follow the corresponding anomalous distribution. The goal is to detect the onset of the dynamic anomaly as quickly as possible, subject to constraints on the frequency of false alarms. This detection problem is posed in a quickest change detection framework where candidate stopping procedures are evaluated according to a delay metric that considers the worst trajectory of the dynamic anomaly. A detection rule is proposed and established to be asymptotically optimal as the mean time to false alarm goes to infinity. Finally, numerical results are provided to validate our theoretical analysis.**

I. Introduction

In quickest change detection (QCD) [1]–[3], the goal is to design stopping procedures to detect a change in the distribution of sequentially observed processes as quickly as possible, subject to *false alarm* (FA) constraints. In the classical single-sensor QCD setting, observations are initially *independent and identically distributed* (i.i.d.) according to a known *non-anomalous* distribution. After some unknown time instant, refered to as the *changepoint*, the observations are generated i.i.d. according to a known *anomalous* distribution. This i.i.d. model has been extensively studied in the QCD literature under two standard formulations: i) the *minimax* setting [4]–[7], where the changepoint is modeled as a deterministic but unknown parameter and the goal is to minimize a worst-case average detection delay subject to a lower bound on the mean time to false alarm; and ii) the *Bayesian* setting [8], [9], where the changepoint is assumed to be a random variable with a known distribution, and the goal is to minimize the average detection delay, subject to a bound on the probability of false alarm.

The theory of QCD has been extensively used to tackle sequential detection problems in the context of sensor networks. In such settings, different QCD problem instances arise according to how the sensors are affected by the anomaly. The simplest setting corresponds to the case where a fixed set of sensors known to the decision maker is affected by the anomaly. In this case, the algorithms proposed in [4]–[9] can be directly applied. A more complicated problem setting arises if we assume that the set of affected sensors is unknown. This case has been extensively studied under the minimax setting and algorithms that are asymptotically optimal have been proposed [10]–[16]. A generalization of the above settings that considers the case that sensors have their data generating distributions altered at different time instants can be tackled by using the detection procedures in [17]–[25]. Note that in all the aforementioned problem formulations, the anomaly eventually affects a fixed set of nodes persistently.

In the dynamic anomaly setting, the anomaly affects different sets of node as time progresses, and it is assumed that the identities of the affected sensor are unknown. In [26], [27] the problem was studied under the assumption that the size of the anomaly is fixed and the anomaly evolves according to a *discrete time Markov chain*. In [28], the Markov chain assumption was lifted, and a worst-path variation of Lorden's [4] detection delay metric was introduced to account for the lack of a specific model for the evolution of the anomaly. A mixture-Cumulative-Sum (CUSUM) test was shown to be exactly optimal with respect to a trade-off formulation using this worst-path delay metric, when the sensors are *homogeneous*, i.e., when the data-generating distributions are the same across sensors before and after the changepoint. In [29], the problem of detecting a dynamic anomaly of growing size was studied and an asymptotically optimal test was constructed for this setting.

In this paper, we extend on the results of [28] to the case of *heterogeneous* sensors, where the non-anomalous and anomalous data generating distributions could be different across the sensors. We focus on the case of an anomaly of size one. The results can be extended to arbitrary anomaly size but the analysis becomes more detailed. We show that an appropriately weighted mixture-CUSUM test is asymptotically optimum in this setting. We finish by comparing the performance of our test to mixture-CUSUM tests that do not use the optimal choice of weights via numerical simulations.

II. Problem Model

Consider a sensor network comprised of L nodes, and let $[L] \triangleq \{1, \dots, L\}$. Denote by $\{\boldsymbol{X}[k]\}_{k=1}^{\infty}$ the sequence of observations that is sequentially sampled by the centralized decision maker. We define by $\boldsymbol{X}[k] = [X_1[k], \ldots, X_L[k]]^\top$ the observation vector at time k, where $X_\ell[k]$ denotes the measurement obtained by sensor $\ell \in [L]$ at time k.

Define by $g_\ell(\cdot)$ and $f_\ell(\cdot)$ the non-anomalous and anomalous *probability density functions* (pdfs), respectively, at sensor ℓ. Initially, the observed process is generated according to the non-anomalous distribution at each sensor, until some *unknown* but *deterministic* changepoint $\nu \geq 0$. In addition, it is assumed that the observations are independent across sensors, i.e., when $k \leq \nu$ the joint pdf of $\boldsymbol{X}[k]$ is given by

$$g(\boldsymbol{X}[k]) \triangleq \prod_{\ell=1}^{L} g_\ell(X_\ell[k]). \tag{1}$$

After the changepoint, the system enters an abnormal state because of the emergence of a dynamic anomaly. As a result one node, which may be different at each time instant, is affected by the anomaly. While anomalous, the observations at the sensor are generated according to the corresponding anomalous pdf. Define the process $S \triangleq \{S[k]\}_{k=1}^{\infty}$ where $S[k]$ contains the index of the node affected by the anomaly at time k. For notational convenience, $S[k]$ is defined for all $k \geq 1$ and not simply for $k > \nu$. When $k > \nu$, we then have that the joint pdf of $\boldsymbol{X}[k]$ *conditioned* on S is given by

$$p_{S[k]}(\boldsymbol{X}[k]) \triangleq f_{S[k]}(X_{S[k]}[k]) \cdot \left(\prod_{\ell \notin S[k]} g_\ell(X_\ell[k]) \right), \tag{2}$$

where $p_\ell(\cdot)$ for $\ell \in [L]$ denotes the joint pdf of a vector observation at a time instant when the underlying distribution is the one induced when node ℓ is the anomalous node. As a result, conditioned on S we have that the $\boldsymbol{X}[k]$'s are independent conditioned on the changepoint, and

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]) & 1 \leq k \leq \nu \\ p_{S[k]}(\boldsymbol{X}[k]) & k > \nu. \end{cases} \tag{3}$$

In this paper, our goal is to design detection procedures in the form of stopping times [1]–[3] that detect the change in distribution described in (3), while offering strong theoretical guarantees. In particular, define by $\mathbb{E}_\infty[\cdot]$ the expectation when no anomaly is present. To quantify the frequency of FA events we use the *mean time to false alarm* (MTFA), denoted by $\mathbb{E}_\infty[\tau]$, for stopping time τ. Furthermore, since the path of the anomaly S is assumed to be deterministic, we use a modification of Lorden's delay metric [4] to account for the worst-path of the anomaly. In particular, define by $\mathbb{E}_\nu^S[\cdot]$ the expectation when changepoint is at time ν, and the trajectory of the anomaly is completely specified by S. Denote by $\mathscr{F}_k = \sigma(\boldsymbol{X}[1], \ldots, \boldsymbol{X}[k])$ the σ-algebra generated by $\boldsymbol{X}[1], \ldots, \boldsymbol{X}[k]$. For any stopping rule τ adapted to the filtration $\mathscr{F} \triangleq \{\mathscr{F}_k\}_{k=1}^{\infty}$, we define the detection delay

$$\mathrm{WADD}(\tau) = \sup_{S} \sup_{\nu \geq 0} \operatorname{ess\,sup} \mathbb{E}_\nu^S \left[\tau - \nu | \tau > \nu, \mathscr{F}_\nu\right], \tag{4}$$

where the convention that $\mathbb{E}_\nu^S \left[\tau - \nu | \tau > \nu, \mathscr{F}_\nu\right] \triangleq 1$ when $\mathbb{P}_\nu^S(\tau > \nu) = 0$ is used. For $\gamma > 1$ a pre-determined constant, define the class of stopping times

$$\mathcal{C}_\gamma \triangleq \{\tau : \mathbb{E}_\infty[\tau] \geq \gamma\}. \tag{5}$$

Our goal is to design τ to solve the following problem:

$$\begin{aligned} \min_{\tau} \quad & \mathrm{WADD}(\tau) \\ \text{s.t.} \quad & \tau \in \mathcal{C}_\gamma. \end{aligned} \tag{6}$$

In addition to the statistical model in (3), for the purposes of performance analysis as we will see in Sections III, IV and V, we also introduce the model that arises when the anomalous node is chosen at random. In particular, define the *probability mass function* (pmf) $\boldsymbol{\alpha} = \{\alpha_\ell : \ell \in [L]\} \in \mathcal{A}$ containing the probabilities that each of the different sensors is chosen to be anomalous. Here, $\mathcal{A}$ denotes the simplex of all probability vectors of dimension L. By placing the anomalous node at random according to $\boldsymbol{\alpha}$ at each time instant after the changepoint, we have that the induced pdf is given by

$$\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[k]) \triangleq \sum_{\ell=1}^{L} \alpha_\ell p_\ell(\boldsymbol{X}[k]). \tag{7}$$

As a result, when the anomaly is placed at random according to $\boldsymbol{\alpha}$ we have the following observation model:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]) & 1 \leq k \leq \nu \\ \overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[k]) & k > \nu. \end{cases} \tag{8}$$

Note that in the observation model of (8), the non-anomalous and anomalous pdfs are completely specified, and hence the underlying QCD problem falls into the classical QCD setting studied in [4]–[7]. For stopping time τ, define the detection delay corresponding to the QCD problem defined in (8) by

$$\overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau) = \sup_{\nu \geq 0} \operatorname{ess\,sup} \overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}[\tau - \nu | \tau > \nu, \mathscr{F}_\nu] \tag{9}$$

where $\overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}[\cdot]$ denotes the expectation when the underlying statistical model is that of (8) and the changepoint is ν. Here, we also use the convention that $\overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}[\tau - \nu | \tau > \nu, \mathscr{F}_\nu] \triangleq 1$ when $\overline{\mathbb{P}}_\nu^{\boldsymbol{\alpha}}(\tau > \nu) = 0$.

III. Universal Lower Bound on the WADD

We begin our analysis by deriving an asymptotic lower bound on WADD for any τ satisfying the false alarm constraint $\mathbb{E}_\infty[\tau] \geq \gamma$. For fixed $\boldsymbol{\alpha}$, define the Kullback-Leibler (KL) number corresponding to the statistical model in (8) by

$$I_{\boldsymbol{\alpha}} \triangleq \overline{\mathbb{E}}_0^{\boldsymbol{\alpha}} \left[\log \frac{\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right]. \tag{10}$$

We then have the following asymptotic lower bound:

Theorem 1. *Let* $\boldsymbol{\alpha}^* \triangleq \arg\min_{\boldsymbol{\alpha} \in \mathcal{A}} I_{\boldsymbol{\alpha}}$ *be the pmf that minimizes the KL number defined in* (10). *We then have that*

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \geq \frac{\log \gamma}{I_{\boldsymbol{\alpha}^*}} (1 + o(1)) \tag{11}$$

as $\gamma \to \infty$.

Proof. By following a similar analysis to that in Theorem 1 of [28], it can be shown that for any $\boldsymbol{\alpha} \in \mathcal{A}$ and $\gamma > 1$

$$\inf_{\tau \in \mathcal{C}_\gamma} \text{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \overline{\text{WADD}}_{\boldsymbol{\alpha}}(\tau). \tag{12}$$

Since (12) holds for all $\boldsymbol{\alpha}$ it also holds for $\boldsymbol{\alpha}^*$, hence

$$\inf_{\tau \in \mathcal{C}_\gamma} \text{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \overline{\text{WADD}}_{\boldsymbol{\alpha}^*}(\tau) \sim \frac{\log \gamma}{I_{\boldsymbol{\alpha}^*}} \tag{13}$$

where $f(x) \sim g(x)$ is used to denote that $g(x) = f(x)(1 + o(1))$ as $x \to \infty$, and the asymptotic approximation follows from the asymptotic analysis of the CUSUM test [4], [7]. □

IV. Minimizing the Kullback-Leibler Number

In this section, we study the minimization of $I_{\boldsymbol{\alpha}}$. We show that $\boldsymbol{\alpha}^*$ is an interior point of $\mathcal{A}$ and establish an important property of the likelihood ratio corresponding to (8) that is going to be crucial in deriving our asymptotically optimal stopping procedure.

Theorem 2. *Let $\boldsymbol{\alpha}^*$ be defined as in Theorem 1. We then have that $\boldsymbol{\alpha}^*$ is an interior point of $\mathcal{A}$ and that*

$$\mathbb{E}_{p_\ell}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = \mathbb{E}_{p_{\ell'}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \tag{14}$$

for all $\ell, \ell' \in [L]$, where $\mathbb{E}_{p_\ell}[\cdot]$ denotes the expectation when ℓ is anomalous.

Proof. Define $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{L-1}]^\top$ where $\alpha_\ell \triangleq \beta_\ell$ for $\ell \in [L-1]$. The constrained optimization of $I_{\boldsymbol{\alpha}}$ can then be equivalently replaced by

$$\begin{aligned} \inf_{\boldsymbol{\beta}} \quad & q(\boldsymbol{\beta}) \\ \text{s.t.} \quad & \beta_\ell \geq 0, \ \forall \ \ell \in [L-1] \\ & \sum_{\ell=1}^{L-1} \beta_\ell \leq 1, \end{aligned} \tag{15}$$

where

$$q(\boldsymbol{\beta}) \triangleq \int_{\mathbb{R}^L} \left(\left(1 - \sum_{\ell=1}^{L-1} \beta_\ell\right) p_L(\boldsymbol{x}) + \sum_{\ell=1}^{L-1} \beta_\ell p_\ell(\boldsymbol{x})\right) \log\left(\frac{\left(\left(1 - \sum_{\ell=1}^{L-1} \beta_\ell\right) p_L(\boldsymbol{x}) + \sum_{\ell=1}^{L-1} \beta_\ell p_\ell(\boldsymbol{x})\right)}{g(\boldsymbol{x})}\right) d\boldsymbol{x}. \tag{16}$$

We then have that for all $\ell \in [L-1]$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_\ell}\right|_{\boldsymbol{\beta}^*} = \mathbb{E}_{p_\ell}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] - \mathbb{E}_{p_L}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right]. \tag{17}$$

Assume that $\boldsymbol{\beta}^*$, the solution of (15), is not an interior point solution. Without loss of generality, this implies that $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_\eta^*, 0, \ldots, 0]^\top$, where $\eta \in [L-2]$ and $0 < \beta_\ell^* < 1$ for all $\ell \in [\eta]$ or $\boldsymbol{\beta}^* = [0 \cdots 0]$. For the first case, we have that for $\ell \in [\eta]$,

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_\ell}\right|_{\boldsymbol{\beta}^*} = 0, \tag{18}$$

which implies that for all $\ell \in [\eta]$

$$\mathbb{E}_{p_\ell}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = \mathbb{E}_{p_L}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \triangleq J. \tag{19}$$

Furthermore, we have that since $\alpha_\ell^* = 0$ for $\eta < \ell < L$

$$\begin{aligned} J &= \left(\sum_{j=1}^{\eta} \beta_j^* + \left(1 - \sum_{j=1}^{\eta} \beta_j^*\right)\right) J = \left(\sum_{j=1}^{\eta} \alpha_j^* + \alpha_L^*\right) J \\ &= \sum_{j=1}^{L} \alpha_j^* \mathbb{E}_{p_j}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = \mathbb{E}_{\overline{p}_{\boldsymbol{\alpha}^*}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \\ &= I_{\boldsymbol{\alpha}^*} > 0. \end{aligned} \tag{20}$$

In addition, for $\eta < \ell < L$, we have that

$$\begin{aligned} &\mathbb{E}_{p_\ell}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \\ &= \mathbb{E}_{p_\ell}\left[\log\left(\sum_{j=1}^{\eta} \alpha_j^* \frac{f_j(X_j)}{g_j(X_j)} + \alpha_L^* \frac{f_L(X_L)}{g_L(X_L)}\right)\right] \\ &= \mathbb{E}_{g}\left[\log\left(\sum_{j=1}^{\eta} \alpha_j^* \frac{f_j(X_j)}{g_j(X_j)} + \alpha_L^* \frac{f_L(X_L)}{g_L(X_L)}\right)\right] \\ &= \mathbb{E}_{g}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] < 0. \end{aligned} \tag{21}$$

We then have that from eqs. (17), (19) - (21)

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_\ell}\right|_{\boldsymbol{\beta}^*} < 0 \tag{22}$$

for all $\eta < \ell < L$, which leads to a contradiction, since (22) cannot hold at the minimum. As a result, $\boldsymbol{\beta}^*$ cannot be of the form $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_\eta^*, 0, \ldots, 0]^\top$, with $0 < \beta_\ell^* < 1$ for $\ell \in [\eta]$. Now assume that $\boldsymbol{\beta}^* = [0 \cdots 0]^\top$. This implies that for all $\ell \in [L-1]$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_\ell}\right|_{\boldsymbol{\beta}^*} = \mathbb{E}_{g_L}\left[\log\left(\frac{f_L(X_L)}{g_L(X_L)}\right)\right] - \mathbb{E}_{f_L}\left[\log\left(\frac{f_L(X_L)}{g_L(X_L)}\right)\right] < 0, \tag{23}$$

which also leads to a contradiction. As a result, $\boldsymbol{\beta}^*$ has to be an interior point solution, which also implies that $\boldsymbol{\alpha}^*$ is an interior point of $\mathcal{A}$. We then have that for all $\ell \in [L-1]$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_\ell}\right|_{\boldsymbol{\beta}^*} = 0 \tag{24}$$

which implies that for all $\ell \in [L-1]$

$$\mathbb{E}_{p_\ell}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = \mathbb{E}_{p_L}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \tag{25}$$

establishing the theorem. □

V. Asymptotically Optimal Test

Define the test statistic of the *mixture-CUSUM* test [28] corresponding to the QCD problem of (8), (9) when $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ by

$$W[k] \triangleq \max_{1 \le i \le k} \sum_{j=i}^{k} Z[j] \tag{26}$$

where

$$Z[j] \triangleq \log \frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])} \tag{27}$$

the log-likelihood ratio at time j and

$$\tau_C = \inf\{k \ge 1 : W[k] \ge b\} \tag{28}$$

the corresponding stopping time. The test statistic in (26) can be expressed recursively as

$$W[k] = (W[k-1])^+ + Z[k] \tag{29}$$

where $W[0] \triangleq 0$ and $(x)^+ = \max\{x, 0\}$.

The intuition behind the test is based on using Theorem 2 to design an equalizer rule with respect to the placement of the anomaly. In particular, from Theorem 2, the expected drift of the statistic in (26) is independent of S.

Next, we use Theorems 1 and 2 to establish that the test defined in eqs. (26) - (29) is asymptotically optimal. In particular, we have the following theorem.

Theorem 3. *Consider the mixture-CUSUM test defined in* (26) - (29). *Assume that*

$$\max_{j\in[L]} \mathbb{E}_{p_j}\left[\left(\log \frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)^2\right] < \infty \tag{30}$$

By choosing $b = \log \gamma$ *we have that as* $\gamma \to \infty$

$$\inf_{\tau\in\mathcal{C}_\gamma} \mathrm{WADD}(\tau) \sim \mathrm{WADD}(\tau_C) \sim \frac{\log\gamma}{I_{\boldsymbol{\alpha}^*}}. \tag{31}$$

Proof. We begin by upper bounding the delay of the test for threshold b by following the proof technique in [7]. Due to the structure of the test we have that

$$\mathrm{WADD}(\tau_C) = \sup_S \mathbb{E}_0^S[\tau_C]. \tag{32}$$

Let $0 < \epsilon < I_{\boldsymbol{\alpha}^*}$ and $n_b = \frac{b}{I_{\boldsymbol{\alpha}^*}-\epsilon}$. Then for any path $S = \{S[k]\}_{k=1}^{\infty}$ from the sum-integral inequality we have that

$$\sup_S \mathbb{E}_0^S\left[\frac{\tau_C}{n_b}\right] = \sup_S \int_0^\infty \mathbb{P}_0^S\left(\frac{\tau_C}{n_b} > x\right)dx \le \sup_S \sum_{\zeta=0}^{\infty} \mathbb{P}_0^S(\tau_C > \zeta n_b) = 1 + \sup_S \sum_{\zeta=1}^{\infty}\mathbb{P}_0^S(\tau_C > \zeta n_b). \tag{33}$$

Then, we have that for any path $S = \{S[k]\}_{k=1}^{\infty}$, $\zeta \ge 1$,

$$\begin{aligned}
\mathbb{P}_0^S(\tau_C > \zeta n_b) &= \mathbb{P}_0^S\left(\max_{1\le k\le \zeta n_b} W[k] < b\right) \\
&= \mathbb{P}_0^S\left(\max_{1\le k\le\zeta n_b}\max_{1\le i\le k}\sum_{j=i}^{k} Z[j] < b\right) \\
&\le \mathbb{P}_0^S\left(\max_{1\le i\le m n_b}\sum_{j=i}^{m n_b} Z[j] < b, \forall\, m\in[\zeta]\right) \\
&\le \mathbb{P}_0^S\left(\sum_{j=(m-1)n_b+1}^{m n_b} Z[j] < b, \forall\, m\in[\zeta]\right) \\
&= \mathbb{P}_0^S\left(\frac{\sum_{j=(m-1)n_b+1}^{m n_b} Z[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon, \forall\, m\in[\zeta]\right) \\
&= \prod_{m=1}^{\zeta}\mathbb{P}_0^S\left(\frac{\sum_{j=(m-1)n_b+1}^{m n_b} Z[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right),
\end{aligned} \tag{34}$$

where the last equality follows due to independence of the observations over time. Note that then for any $b > 0$ we have that

$$\begin{aligned}
&\sup_S \sum_{\zeta=1}^{\infty}\mathbb{P}_0^S(\tau_C > \zeta n_b) = \sup_S \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\mathbb{P}_0^S(\tau_C > \zeta n_b) \\
&\le \lim_{\xi\to\infty}\sup_S\sum_{\zeta=1}^{\xi}\mathbb{P}_0^S(\tau_C > \zeta n_b) \le \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\sup_S\mathbb{P}_0^S(\tau_C > \zeta n_b) \\
&\le \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\sup_S\left[\prod_{m=1}^{\zeta}\mathbb{P}_0^S\left(\frac{\sum_{j=(m-1)n_b+1}^{m n_b} Z[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right)\right] \\
&\le \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\prod_{m=1}^{\zeta}\left[\sup_S\mathbb{P}_0^S\left(\frac{\sum_{j=(m-1)n_b+1}^{m n_b} Z[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right)\right] \\
&= \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\left[\sup_S\mathbb{P}_0^S\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right)\right]^{\zeta}.
\end{aligned} \tag{35}$$

Note that for any path S we also have that

$$\mathbb{P}_0^S\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right) \le \mathbb{P}_0^S\left(\left|\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} - I_{\boldsymbol{\alpha}^*}\right| > \epsilon\right). \tag{36}$$

In addition, from Theorem 2 we have that

$$\mathbb{E}_0^S\left[\frac{\sum_{j=1}^{n_b} Z[j]}{n_b}\right] = \sum_{j=1}^{n_b} \mathbb{E}_{p_{S[j]}}\left[\frac{Z[j]}{n_b}\right] = I_{\boldsymbol{\alpha}^*}. \tag{37}$$

Define

$$\bar{\sigma}^2 \triangleq \max_{j\,\in\,[L]} \mathrm{Var}_{p_j}\left[\log \frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right]. \tag{38}$$

From eq. (30), we have that $\bar{\sigma}^2 < \infty$. Then, by Chebychev's inequality

$$\mathbb{P}_0^S\left(\left|\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} - I_{\boldsymbol{\alpha}^*}\right| > \epsilon\right) \leq \mathrm{Var}_0^S\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b}\right)\frac{1}{\epsilon^2}$$

$$= \frac{1}{\epsilon^2 n_b^2}\sum_{j=1}^{n_b} \mathrm{Var}_{p_{S[j]}}\left(Z[j]\right) \leq \frac{\sum_{j=1}^{n_b} \bar{\sigma}^2}{n_b^2\epsilon^2} = \frac{\bar{\sigma}^2}{n_b\epsilon^2}. \tag{39}$$

By using (33), (35) and (39) we then have that

$$\sup_S \mathbb{E}_0^{\boldsymbol{S}}\left[\frac{\tau_C}{n_b}\right] \leq 1 + \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\left[\frac{\bar{\sigma}^2}{n_b\epsilon^2}\right]^{\zeta}. \tag{40}$$

Let $0 < \delta < 1$. Since n_b is increasing with b, we have that for all $b > B$, where B large enough

$$\sup_S \mathbb{E}_0^S\left[\frac{\tau_C}{n_b}\right] \leq 1 + \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\delta^{\zeta} = \sum_{\zeta=0}^{\infty}\delta^{\zeta} = \frac{1}{1-\delta} \tag{41}$$

which implies that for all $b > B$

$$\sup_S \mathbb{E}_0^S\left[\tau_C\right] \leq \frac{b}{(I_{\boldsymbol{\alpha}^*} - \epsilon)(1-\delta)}. \tag{42}$$

Since (42) holds for all $\epsilon > 0$ we have that

$$\sup_{\boldsymbol{S}} \mathbb{E}_0^{\boldsymbol{S}}\left[\tau_C\right] \leq \frac{b}{I_{\boldsymbol{\alpha}^*}(1-\delta)}. \tag{43}$$

Finally, since $\delta \to 0$ as $b \to \infty$

$$\mathrm{WADD}(\tau_C) = \sup_S \mathbb{E}_0^S\left[\tau_C\right] \leq \frac{b}{I_{\boldsymbol{\alpha}^*}}(1 + o(1)) \tag{44}$$

as $b \to \infty$. By combining Theorem 1 with (44), and since $\mathbb{E}_\infty[\tau_C] \geq \gamma$ is implied when $b = \log\gamma$ for the CUSUM test [4]–[7], the theorem is established. □

VI. Numerical Results

In this section, we conduct numerical simulations for the studied dynamic anomaly QCD problem for the case of $L = 10$ and when $g_\ell = \mathcal{N}(0,1)$ for all $\ell \in [L]$, and $f_\ell = \mathcal{N}(\mu_\ell, 1)$ with $\boldsymbol{\mu} = [1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9]^\top$ denoting the vector of the means of the anomalous pdf. We compare three versions of the test introduced in eqs. (26) - (29): the first version ("Uniform slopes" in Fig. 1) uses the optimal weights $\boldsymbol{\alpha}^*$ to achieve a uniform average statistic drift (which is approximately equal to 0.178) among anomaly placements; the second and third versions ("Non-uniform slopes 1" and "Non-uniform slopes 2" in Fig. 1) use arbitrary choices of weights that only guarantee that the expected drift of the statistic is positive for any placement of the anomaly. In Fig. 1, we see that the mixture-CUSUM test using the optimal weights $\boldsymbol{\alpha}^*$ outperforms the other two implementations. It should be noted that the WADD in this simulation is calculated approximately, since the worst path of the anomaly cannot be specified analytically. However, as the MTFA becomes large, WADD can be approximated by placing the anomaly at only the node that corresponds to the worst post-change expected drift. For the case of "Non-uniform slopes 1" this corresponds to placing the anomaly at sensor 2, and for the case of "Non-uniform slopes 2" at sensor 5. For the optimal weight choice, the placement of the anomaly does not affect the delay for large MTFA, since the expected drift does not depend on the trajectory of the anomaly.

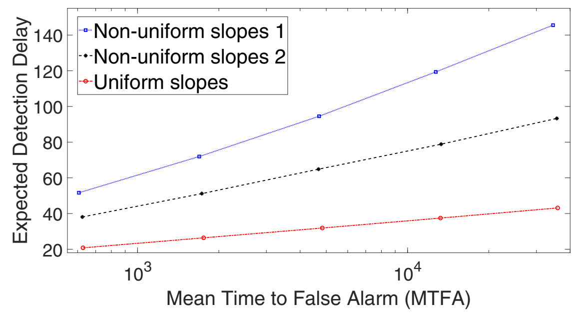

Fig. 1. WADD versus MTFA for test using $\boldsymbol{\alpha}^*$ and tests not using $\boldsymbol{\alpha}^*$ when $L = 10$.

Acknowledgement

This work was supported in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196 (IoBT CRA), and in part by the National Science Foundation (NSF) under grant CCF 16-18658 and CIF 15-14245, through the University of Illinois at Urbana-Champaign, and grant CIF 15-13373, through Rutgers University.

References

[1] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*, Statistics. CRC Press, 2014.

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

[3] V. V. Veeravalli and T. Banerjee, *Quickest Change Detection*, Elsevier: E-reference Signal Processing, 2013.

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

[5] M. Pollak, "Optimal detection of a change in distribution," *Annals of Statistics*, vol. 13, no. 1, pp. 206–227, Mar. 1985.

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

[7] 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, Nov. 1998.

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

[9] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory of Probability & Its Applications*, vol. 49, Jan. 2005.
[10] 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, Jan. 2004.
[11] 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, Sept. 2006.
[12] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, Apr. 2010.
[13] Y. Mei, "Quickest detection in censoring sensor networks," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2011.
[14] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.
[15] G. Fellouris and A. G. Tartakovsky, "Multichannel sequential detection — part I: Non-i.i.d. data," *IEEE Trans. Inform. Theory*, vol. 63, no. 7, pp. 4551–4571, Jul. 2017.
[16] 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.
[17] S. Zou, G. Fellouris, and V. V. Veeravalli, "Quickest change detection under transient dynamics: Theory and asymptotic analysis," *IEEE Trans. Inform. Theory*, 2018.
[18] 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.
[19] 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.
[20] M. Ludkovski, "Bayesian quickest detection in sensor arrays," *Sequential Analysis*, vol. 31, no. 4, pp. 481–504, 2012.
[21] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *arXiv e-prints*, p. arXiv:1807.06143, Jul. 2018.
[22] R. Zhang and Y. Mei, "Asymptotic statistical properties of communication-efficient quickest detection schemes in sensor networks," *Sequential Analysis*, vol. 37, no. 3, pp. 375–396, 2018.
[23] G. V. Moustakides and V. V. Veeravalli, "Sequentially detecting transitory changes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2016.
[24] G. Rovatsos, J. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Statistical power system line outage detection under transient dynamics," *IEEE Trans. Signal Proc.*, vol. 65, no. 11, pp. 2787–2797, Jun. 2017.
[25] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Quickest change detection under transient dynamics," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2017.
[26] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Quickest detection of a moving target in a sensor network," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Paris, France, Jul. 2019.
[27] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Sequential algorithms for moving anomaly detection in networks," *Sequential Analysis*, (to appear).
[28] G. Rovatsos, G. V Moustakides, and V. V. Veeravalli, "Quickest detection of a dynamic anomaly in a sensor network," in *Proc. IEEE 53nd Asilomar Conf. on Signals, Systems, and Computers (ACSSC)*, Pacific Grove, USA, Nov. 2019.
[29] G. Rovatsos, V. V. Veeravalli, D. Towsley, and A. Swami, "Quickest detection of growing dynamic anomalies in networks," *arXiv e-prints*, p. arXiv:1910.09151, Oct. 2019.