Quickest Detection of Moving Anomalies in Sensor Networks

Georgios Rovatsos, *Student Member, IEEE*, George V. Moustakides, *Life Senior Member, IEEE*, and Venugopal V. Veeravalli, *Fellow, IEEE*

***Abstract*—The problem of sequentially detecting a moving anomaly is studied, in which the anomaly affects different parts of a sensor network over time. Each network sensor is characterized by a pre- and post-change distribution. Initially, the observations of each sensor are generated according to the corresponding pre-change distribution. After some *unknown* but *deterministic* time instant, a moving anomaly emerges, affecting different sets of sensors as time progresses. Our goal is to design a stopping procedure to detect the emergence of the anomaly as quickly as possible, subject to *false alarms* constraints. The problem is studied in a quickest change detection framework where it is assumed that the evolution of the anomaly is *unknown* but *deterministic*. A modification of Lorden's detection delay is proposed to account for the trajectory of the anomaly that maximizes the detection delay of a detection procedure. It is established that a *Cumulative Sum*-type test solves the resulting sequential detection problem exactly when the sensors are *homogeneous*. For the case of *heterogeneous* sensors, the proposed detection scheme can be modified to provide a first-order asymptotically optimal algorithm.**

***Index Terms*—Sequential change detection, CUSUM test, dynamic anomaly, worst-path approach, heterogeneous sensors.**

I. Introduction

In QUICKEST change detection (QCD) [3]–[5], a sequentially observed time series undergoes a change in the underlying probability distribution at some unknown time instant. The goal is to design a detection procedure, in the form of a *stopping time*, to detect this abrupt change as quickly as possible, subject to *false alarm* (FA) constraints.

In the case of multisensor networks, the theory of QCD has been widely employed to provide solutions to a variety of detection problems of interest. The simplest case corresponds to the anomaly persistently affecting a fixed set of sensors, the identity of which is known to the decision maker, after the changepoint. This problem is a trivial extension of the classical single-sensor QCD setting, and the algorithms in [6]–[11] can be directly applied to provide performance guarantees. A significantly more complicated problem instance arises if we assume that the decision maker has no knowledge of the identity of the affected nodes. This problem has been extensively studied in the literature under the minimax setting [12]–[18]. Generalizations of these two aforementioned settings consider the case that the onset of the anomaly is perceived at different time instants across sensors [19]–[27]. It is crucial to note that in the sensor network problems studied thus far, a core assumption made is that the anomaly persistently affects each sensor.

In this work, we study the problem of sequentially detecting a moving anomaly under Lorden's minimax framework [6]. In the moving anomaly QCD setting, it is assumed that different sets of nodes are affected by the anomaly as time progresses, and that the anomalous nodes are unknown to the decision maker. As a result, the anomaly does not affect any specific sensor persistently, but is persistent in the network as a whole. The problem was initially studied in [28], [29], where it was assumed that the anomaly evolves according to a *discrete time Markov chain* and is of fixed size. Here we drop the Markov assumption and assume that the trajectory of the anomaly is *unknown* and *deterministic*. To account for the lack of a specific model for the anomaly path, we modify Lorden's detection delay metric [6] to obtain a worst-path detection delay, and frame Lorden's QCD problem with this new delay metric. In the case of a network comprised of *homogeneous* sensors, which share a common pre-change and a common post-change distribution, we establish that a *Cumulative Sum* (CUSUM)-type [30] test that detects a transition to a mixture of distributions, each induced on the observations according to the identity of the anomalous nodes, is *exactly optimal*. Furthermore, we show that in the general case of *heterogeneous* sensors the proposed test can be modified to provide a first-order asymptotically optimal solution.

The problem studied in this work is particularly relevant to applications where the location of the anomalous nodes can change rapidly with time. An example of such a setting is anomaly detection in intrusion detection

Manuscript received October 5, 2020; revised March 10, 2021 and April 19, 2021; accepted April 20, 2021. Date of publication April 27, 2021; date of current version June 21, 2021. This work was supported in part by the Army Research Laboratory through Cooperative Agreement under Grant W911NF-17-2-0196 (IoBT CRA); in part by the National Science Foundation (NSF) under Grant CCF 16-18658 and Grant CIF 15-14245, through the University of Illinois at Urbana-Champaign; and in part by Rutgers University under Grant CIF 15-13373. This work was presented in part in the 2019 Asilomar Conference on Signals, Systems, and Computers [1] and in the 2020 International Symposium on Information Theory [2]. *(Corresponding author: Venugopal V. Veeravalli.)*

Georgios Rovatsos was with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA. He is now with Walmart Labs, Mountain View, CA, USA (e-mail: rovatso2@illinois.edu).

George V. Moustakides is with the Electrical and Computer Engineering Department, University of Patras, 26500 Patra, Greece (e-mail: moustaki@upatras.gr).

Venugopal V. Veeravalli is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA (e-mail: vvv@illinois.edu).

Digital Object Identifier 10.1109/JSAIT.2021.3076043

applications, especially in cases where the intruder's location can change rapidly with time [13], [31]. In such settings, current techniques that assume that the sensors are affected persistently by the anomaly after it appears may incur large detection delays and new algorithms are needed.

II. Problem Model

In this section, we present the statistical model that governs the data generated by the sensor network, and our QCD problem in a delay-FA optimization framework after introducing the worst-trajectory delay metric. We begin by introducing some necessary notation. Our convention in this work is that for any sequence $\{\alpha[k]\}_{k=1}^{\infty}$, if $k_2 > k_1$ we define $\prod_{j=k_2}^{k_1}\alpha[j] \triangleq 1$ and $\sum_{j=k_2}^{k_1}\alpha[j] \triangleq 0$. Furthermore, for any sequence $\{\alpha[k]\}_{k=1}^{\infty}$, $\alpha[k_1,k_2] \triangleq [\alpha[k_1],\ldots\alpha[k_2]]^\top$ denotes the samples from time k_1 to k_2. For a set E, $|E|$ denotes the number of elements in the set. The set $\{1,2,\ldots,K\}$ is denoted by $[K]$; in particular, the set of L senors that comprise the sensor network is denoted by $[L]$. The sequence $\{\boldsymbol{X}[k]\}_{k=1}^{\infty}$ denotes the sequence of observations generated by the sensor network, where $\boldsymbol{X}[k] \triangleq [X_1[k],\ldots,X_L[k]]^\top$ is the observation vector at time k and $X_\ell[k] \in \mathbb{R}$ is the measurement obtained by sensor $\ell \in [L]$ at time k. The filtration generated by the observation process is denoted by $\mathscr{F} \triangleq \{\mathscr{F}_k\}_{k=1}^{\infty}$, where $\mathscr{F}_k = \sigma(\boldsymbol{X}[1,k])$ denotes the σ-algebra generated by $\boldsymbol{X}[1,k]$. Furthermore, for $K \geq 0$, $\|\boldsymbol{x}\|_K$ denotes the l_K norm of vector $\boldsymbol{x}$. Finally, for functions $f:\mathbb{R}\mapsto\mathbb{R}$, $g:\mathbb{R}\mapsto\mathbb{R}$, $f(x)\sim g(x)$ denotes that $g(x)=f(x)(1+o(1))$ as $x\to\infty$, where $o(1)\to 0$ as $x\to\infty$.

A. Observation Model

Denote by $g_\ell(x), f_\ell(x)$ the pre- and post-change *probability density functions* (pdfs) at sensor $\ell \in [L]$, respectively. We assume that at each sensor the corresponding pre- and post-change distributions are different and that all data-generating distributions are known to the decision maker. Initially, the data at all the sensors are i.i.d. according to the pre-change distribution, and observations are assumed to be independent across sensors. As a result, the joint pdf of $\boldsymbol{X}[k]$ is initially given by

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

After some *unknown* and *deterministic changepoint* $\nu \geq 0$, a physical event leads to the emergence of a moving anomaly in the network. The anomaly moves around the network, affecting different sets of size $1 \leq m \leq L$ as time progresses. It is assumed that m is constant and known to the decision maker. Define the process $\boldsymbol{S} \triangleq \{\boldsymbol{S}[k]\}_{k=1}^{\infty}$, where $\boldsymbol{S}[k]$ denotes the m-dimensional vector containing the indices of the anomalous nodes at time k. Note that for notational convenience, $\boldsymbol{S}[k]$ is defined for all $k \geq 1$ and not simply for $k > \nu$. We denote by $\mathcal{E}(L,m) \triangleq \mathcal{E} \triangleq \{\boldsymbol{E}_j \mid 1 \leq j \leq \binom{L}{m}\}$ the set of all distinct possible vector-values that $\boldsymbol{S}[k]$ can take (without loss of generality we assume that the components of each vector are ordered to provide a unique vector per anomaly placement). The nodes affected by the anomaly generate observations according to the post-change mode. In particular, for $k > \nu$, we have that conditioned on $\boldsymbol{S}$, the joint pdf of $\boldsymbol{X}[k]$ is:

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

where for $\boldsymbol{E} \in \mathcal{E}$, $p_{\boldsymbol{E}}(\boldsymbol{x})$ denotes the joint pdf induced on a vector observation when the anomalous nodes are the ones contained in $\boldsymbol{E}$. We also assume that the observations are independent across time, conditioned on the changepoint. As a result, conditioned on ν and $\boldsymbol{S}$ the complete statistical model is the following:

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

Note that this moving anomaly QCD problem can also be posed as the following dynamic composite hypothesis testing problem: at each time instant k, decide between the hypotheses

$$\begin{aligned} &H_{1,\boldsymbol{S}}^k : \nu < k \text{ and anomaly evolves according to } \boldsymbol{S} \\ &H_0^k : \nu \geq k. \end{aligned} \tag{4}$$

The likelihood ratio between the hypothesis that the anomaly appears at time $\nu+1$ and evolves according to $\boldsymbol{S}$ and the hypothesis that the anomaly never appears is given by

$$\Gamma_{\boldsymbol{S}}(k,\nu) \triangleq \prod_{j=\nu+1}^{k}\left(\prod_{\ell\in\boldsymbol{S}[j]}\frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])}\right). \tag{5}$$

B. Delay-FA Trade-off Formulation

In this work, the goal is to design a detection procedure in the form of a *stopping time* to detect the abrupt change in distribution detailed in (3). A stopping time τ [3]–[5] adapted to $\mathscr{F}$ is a positive random variable that satisfies $\{\tau \leq k\} \in \mathscr{F}_k$ for all $k \geq 1$. An efficient stopping procedure offers quick detection while guaranteeing a sufficiently low frequency of false alarms. To frame this trade-off mathematically, we employ a modified version of Lorden's delay-FA formulation [6]. In particular, since the anomaly trajectory process $\boldsymbol{S}$ is assumed to be unknown, we modify Lorden's delay metric to evaluate candidate detection schemes according to the anomaly path that maximizes the expected detection delay. In particular, denote by $\mathbb{E}_\nu^{\boldsymbol{S}}[\,\cdot\,]$ the expectation when the changepoint is equal to ν and the trajectory of the anomaly is specified by $\boldsymbol{S}$. Then, for any stopping rule τ adapted to $\mathscr{F}$ consider the following modification of Lorden's worst average detection delay (WADD) metric:

$$\mathrm{WADD}(\tau) = \sup_{\boldsymbol{S}}\sup_{\nu\geq 0}\operatorname{ess\,sup}\mathbb{E}_\nu^{\boldsymbol{S}}\big[\tau-\nu|\tau>\nu,\mathscr{F}_\nu\big] \tag{6}$$

where the convention that $\mathbb{E}_\nu^{\boldsymbol{S}}[\tau-\nu|\tau>\nu,\mathscr{F}_\nu] \triangleq 1$ when $\mathbb{P}_\nu^{\boldsymbol{S}}(\tau>\nu)=0$ is used. Note that an additional sup is used to account for the trajectory of the anomaly that maximizes the detection delay of τ. Denote 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. For $\gamma > 1$ a pre-determined constant, define the class of stopping times

$$\mathcal{C}_\gamma \triangleq \{\tau : \mathbb{E}_\infty[\tau] \geq \gamma\}. \tag{7}$$

Our goal then is to design a stopping time τ to solve the following problem:

$$\begin{aligned} \min_{\tau} \ & \mathrm{WADD}(\tau) \\ \text{s.t.} \ & \tau \in \mathcal{C}_{\gamma}. \end{aligned} \tag{8}$$

Remark 1: We will without loss of generality be considering stopping times τ satisfying $\mathbb{E}_{\infty}[\tau] < \infty$, since any stopping time that does not satisfy this condition can be truncated to provide a smaller detection delay while at the same time satisfying the FA constraint.

C. Randomized Anomaly Allocation Model

Before proceeding to the presentation of our main theoretical results, we introduce another statistical model that plays an important role in the mathematical analysis, as well as in the interpretation of the results. In particular, consider an alternate setting to that of (3), where at each time instant after the changepoint, the m anomalous nodes are chosen *randomly*. To this end, denote by $\boldsymbol{\alpha} = \{\alpha_{\boldsymbol{E}} : \boldsymbol{E} \in \mathcal{E}\} \in \mathcal{A}$ the *probability mass function* (pmf) containing the probabilities that each of the vectors in $\mathcal{E}$ is chosen as the vector of anomalous nodes. That is, at each time instant k the probability that the m anomalous nodes are chosen to be in $\boldsymbol{E}$ is given by $\alpha_{\boldsymbol{E}}$, and the set of anomalous nodes is chosen i.i.d. across time. Here, $\mathcal{A}$ denotes the simplex of all probability vectors of dimension $|\mathcal{E}|$. When at each time instant after the changepoint the anomalous nodes are placed i.i.d. randomly according to $\boldsymbol{\alpha}$, we have that the induced pdf after the changepoint is the mixture:

$$\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[k]) \triangleq \sum_{\boldsymbol{E} \in \mathcal{E}} \alpha_{\boldsymbol{E}} p_{\boldsymbol{E}}(\boldsymbol{X}[k]). \tag{9}$$

As a result, the complete observation model for the case of a randomized anomaly allocation according to pmf $\boldsymbol{\alpha}$ is:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]) & 1 \leq k \leq \nu \\ \overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[k]) & k > \nu. \end{cases} \tag{10}$$

Similarly to (3), we can pose the following dynamic composite hypothesis testing problem corresponding to (10): at each time k choose between the hypotheses

$\bar{H}^{k}_{1,\boldsymbol{\alpha}} : \nu < k$ and anomaly placed randomly according to $\boldsymbol{\alpha}$
$\bar{H}^{k}_{0} : \nu \geq k.$

The likelihood ratio between the hypothesis that the anomaly appears at time $\nu + 1$ and is randomly placed according to $\boldsymbol{\alpha}$ at each time instant and the hypothesis that the anomaly never appears is given by

$$\begin{aligned} \mathcal{L}_{\boldsymbol{\alpha}}(k, \nu) &\triangleq \prod_{j=\nu+1}^{k} \frac{\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])} \\ &= \prod_{j=\nu+1}^{k} \left(\sum_{\boldsymbol{E} \in \mathcal{E}} \alpha_{\boldsymbol{E}} \prod_{\ell \in \boldsymbol{E}} \frac{f_{\ell}(X_{\ell}[j])}{g_{\ell}(X_{\ell}[j])} \right) \\ &= \prod_{j=\nu}^{k} \mathcal{L}_{\boldsymbol{\alpha}}(j, j-1). \end{aligned} \tag{11}$$

The Kullback-Leibler (KL) divergence between the post- and pre-change distributions in (10) given by

$$I_{\boldsymbol{\alpha}} \triangleq \overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{0} \left[\log \frac{\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right], \tag{12}$$

where $\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu}[\cdot]$ denotes the expectation when the underlying statistical model is that of (10) with changepoint being equal to ν and the anomaly placed randomly according to $\boldsymbol{\alpha}$.

This QCD problem is associated with a corresponding detection delay. In particular, for stopping time τ, define the detection delay corresponding to the model in (10) by

$$\overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau) = \sup_{\nu \geq 0} \operatorname{ess\,sup} \overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu} \left[\tau - \nu | \tau > \nu, \mathscr{F}_{\nu} \right]. \tag{13}$$

Here, we also use the convention that $\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu}[\tau - \nu | \tau > \nu, \mathscr{F}_{\nu}] \triangleq 1$ when $\overline{\mathbb{P}}^{\boldsymbol{\alpha}}_{\nu}(\tau > \nu) = 0$. Since both the pre- and post-change joint pdfs for the QCD problem presented in (10)–(13) are completely specified, the classical CUSUM test studied in [6]–[9] can be directly applied to solve this QCD problem exactly [8]. In the remainder of the paper, we show that solving the QCD problem in (10)–(13) for a specific choice of $\boldsymbol{\alpha}$, which depends in the data generating distributions of the sensors, can lead to a solution to the QCD problem described in (3)–(8).

III. Proposed Detection Algorithm

For $\boldsymbol{\lambda} \in \mathcal{A}$, consider the following Mixture-CUSUM (M-CUSUM) test statistic

$$W_{\boldsymbol{\lambda}}[k] \triangleq \max_{1 \leq i \leq k} \mathcal{L}_{\boldsymbol{\lambda}}(k, i-1) \tag{14}$$

with the corresponding stopping time

$$\tau_{W}(\boldsymbol{\lambda}, b) \triangleq \inf \left\{ k \geq 1 : W_{\boldsymbol{\lambda}}[k] \geq e^{b} \right\} \tag{15}$$

where $b > 0$ is a constant chosen so that the stopping time satisfies the FA constraint in (7). The test statistic in (14) can be computed recursively as

$$W_{\boldsymbol{\lambda}}[k] = \max\{W_{\boldsymbol{\lambda}}[k-1], 1\} \mathcal{L}_{\boldsymbol{\lambda}}(k, k-1) \tag{16}$$

where $W_{\boldsymbol{\lambda}}[0] \triangleq 0$.

Remark 2: Note that according to (12) and (17) the calculation of the M-CUSUM test statistic at each time instant requires the calculation of $\binom{L}{m}$ terms inside the log-likelihood ratio. In practice an approximation of the log-likelihood ratio function (e.g., discretized version) can be stored and used to calculate the test statistic significantly more efficiently.

Note that the M-CUSUM test presented in (14)–(16) is the exact solution to the QCD problem detailed in (10)–(13) when $\boldsymbol{\alpha} = \boldsymbol{\lambda}$, if b is chosen such that $\mathbb{E}_{\infty}[\tau_{W}(\boldsymbol{\lambda}, b)] = \gamma$ [8]. In the remainder of the paper, we establish that by choosing $\boldsymbol{\lambda}$ accordingly the M-CUSUM procedure is also an exact solution to (8) when the network is comprised of homogeneous sensors, as well as first-order asymptotically optimal for the general heterogeneous network case. Our analysis is based on relating the two QCD models presented in Section II and exploiting tools used for the analysis of the CUSUM test in [8], [9]. We begin by presenting an important theorem relating the detection delay metrics (6), (13) introduced in Section II.

Theorem 1: Let $\gamma > 1$ and $\boldsymbol{\alpha} \in \mathcal{A}$. For the M-CUSUM test introduced in (14)-(16) with b chosen such that $\mathbb{E}_\infty[\tau_W(\boldsymbol{\alpha}, b)] = \gamma$ we have that

$$\begin{aligned} \mathrm{WADD}(\tau_W(\boldsymbol{\alpha}, b)) &\geq \inf_{\tau \in C_\gamma} \mathrm{WADD}(\tau) \\ &\geq \overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau_W(\boldsymbol{\alpha}, b)). \end{aligned} \quad (17)$$

Proof: See the Appendix. ■

IV. Homogeneous Sensor Network Case

In this section, we consider the case of a homogeneous sensor network, i.e., a network where $g_\ell(x) \triangleq g(x)$ and $f_\ell(x) \triangleq f(x)$ for all $\ell \in [L]$, $x \in \mathbb{R}$ (note that with some abuse of notation $g(x)$ denotes the common marginal pre-change pdf, while $g(\boldsymbol{x})$ denotes the joint pdf under $\mathbb{P}_\infty(\cdot)$). Since the network is symmetric, an intuitive weight choice is one where all the weights in the M-CUSUM test of Section III test to be equal. This then implies by symmetry arguments that placing the anomaly randomly or with the worst-path approach will not lead to a different detection delay. In particular, we have the following lemma:

Lemma 1: Consider a homogeneous sensor network where $g_\ell(x) \triangleq g(x)$ and $f_\ell(x) \triangleq f(x)$ for all $\ell \in [L]$, $x \in \mathbb{R}$. Let $\boldsymbol{\lambda}_U \triangleq \left[\frac{1}{\binom{L}{m}}, \ldots, \frac{1}{\binom{L}{m}}\right]^\top$ be the uniform M-CUSUM weights vector. For any threshold $b > 0$ and any $\boldsymbol{\alpha} \in \mathcal{A}$ we have that

$$\mathrm{WADD}(\tau_W(\boldsymbol{\lambda}_U, b)) = \overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau_W(\boldsymbol{\lambda}_U, b)). \quad (18)$$

Proof: See the Appendix. ■

By using Theorem 1 and Lemma 1 we can establish the exact optimality of the M-CUSUM test with uniform weights for the case of a homogeneous sensor network.

Theorem 2: Consider a homogeneous sensor network where $g_\ell(x) \triangleq g(x)$ and $f_\ell(x) \triangleq f(x)$ for all $\ell \in [L]$, $x \in \mathbb{R}$. Let $\gamma > 1$. The M-CUSUM test with uniform weights $\boldsymbol{\lambda} = \boldsymbol{\lambda}_U \triangleq \left[\frac{1}{\binom{L}{m}}, \ldots, \frac{1}{\binom{L}{m}}\right]^\top$ and threshold b chosen such that $\mathbb{E}_\infty[\tau_W(\boldsymbol{\lambda}_U, b)] = \gamma$ is exactly optimal with respect to (8):

$$\mathrm{WADD}(\tau_W(\boldsymbol{\lambda}_U, b)) = \inf_{\tau \in C_\gamma} \mathrm{WADD}(\tau). \quad (19)$$

Proof: The result follows directly by combining Theorem 1 and Lemma 1. ■

Theorem 2 implies that, for the case of homogeneous sensors, the M-CUSUM test that solves the QCD problem of (10)–(13) for a uniform pmf $\boldsymbol{\alpha} = \boldsymbol{\lambda}_U$ is also the exact solution to (3)–(8). Next, we investigate whether a similar result holds for the general case of heterogeneous networks.

V. Heterogeneous Sensor Network Case

In Section IV, we saw how the symmetry of a homogeneous sensor network can facilitate the construction of an exactly optimal test with respect to (8). However, in the case of a heterogeneous sensor network, such a symmetry no longer holds, and a result similar to Lemma 1 cannot be established in general. In this section, we show that by choosing the weights of the M-CUSUM test carefully, a first-order asymptotically optimal test can be derived.

A. Universal Asymptotic Lower Bound on the WADD

We begin our analysis by presenting an asymptotic lower bound on WADD for stopping times satisfying the false alarm constraint $\mathbb{E}_\infty[\tau] \geq \gamma$. Our lower bound is derived by using Theorem 1 together with the asymptotic lower bound on $\overline{\mathrm{WADD}}$ [6], [9]. In particular, note that the inequalities in Theorem 1 hold for any arbitrary $\boldsymbol{\alpha} \in \mathcal{A}$. Therefore, to obtain the tightest asymptotic lower bound we need to consider the $\boldsymbol{\alpha}$ that maximizes the coefficient of the asymptotic rate of $\overline{\mathrm{WADD}}$. To this end, define the minimizer of the effective KL divergence [32] $I_{\boldsymbol{\alpha}}$ by

$$\boldsymbol{\alpha}^* \triangleq \arg\min_{\boldsymbol{\alpha} \in \mathcal{A}} I_{\boldsymbol{\alpha}}. \quad (20)$$

It can be shown that $I_{\boldsymbol{\alpha}}$ is strictly convex with respect to $\boldsymbol{\alpha}$, hence, such a minimizer is uniquely defined. We then have the following theorem:

Theorem 3: Let $\boldsymbol{\alpha}^*$ be defined as in (20). Then

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \geq \frac{\log \gamma}{I_{\boldsymbol{\alpha}^*}}(1 + o(1)) \quad (21)$$

as $\gamma \to \infty$.

Proof: By Theorem 1 we have that for any $\boldsymbol{\alpha} \in \mathcal{A}$ and any $\gamma > 1$

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau). \quad (22)$$

which implies that the inequality also holds for $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, i.e.,

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \overline{\mathrm{WADD}}_{\boldsymbol{\alpha}^*}(\tau) \sim \frac{\log \gamma}{I_{\boldsymbol{\alpha}^*}} \quad (23)$$

where the asymptotic delay approximation follows from the asymptotic analysis of the CUSUM test [6], [9]. ■

B. Asymptotic Upper Bound on the WADD of M-CUSUM Test

Although deriving a lower bound on WADD is similar for both homogeneous and heterogeneous sensor networks (Theorem 1), upper bounding WADD in the latter case for arbitrary $\boldsymbol{\lambda}$ is nontrivial. To this end, we present the following lemma:

Lemma 2: Let $\boldsymbol{\alpha}^*$ be defined as in (20). We then have the following:

i) Case $m \geq 2$: $\boldsymbol{\alpha}^*$ cannot be a corner point of $\mathcal{A}$, i.e., $2 \leq \|\boldsymbol{\alpha}^*\|_0 \leq |\mathcal{E}|$.

If $\|\boldsymbol{\alpha}^*\|_0 = |\mathcal{E}|$ (interior-point minimum),

$$\mathbb{E}_{p_{\boldsymbol{E}}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = \mathbb{E}_{p_{\boldsymbol{E}'}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \quad (24)$$

for all $\boldsymbol{E}$, $\boldsymbol{E}' \in \mathcal{E}$, where $\mathbb{E}_{p_{\boldsymbol{E}}}[\,\cdot\,]$ denotes the expected value when the set of anomalous nodes is given by $\boldsymbol{E} \in \mathcal{E}$.

If $2 \leq \|\boldsymbol{\alpha}^*\|_0 < |\mathcal{E}|$ (boundary-point minimum), let $\mathcal{E}' \triangleq \{\boldsymbol{E} \in \mathcal{E} : \alpha^*_{\boldsymbol{E}} > 0\}$ the subset of vectors in $\mathcal{E}$ for which non-zero weights are assigned in $\boldsymbol{\alpha}^*$. We then have that for all $\boldsymbol{E}$, $\boldsymbol{E}' \in \mathcal{E}'$ (24) holds. Furthermore, we have that for all $\boldsymbol{B} \in \mathcal{E}'$, $\boldsymbol{B}' \in \mathcal{E} \setminus \mathcal{E}'$

$$\mathbb{E}_{p_{\boldsymbol{B}'}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] > \mathbb{E}_{p_{\boldsymbol{B}}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right]. \quad (25)$$

ii) Case $m = 1$ (single anomalous node): $\boldsymbol{\alpha}^*$ is an interior point of $\mathcal{A}$, i.e., $\|\boldsymbol{\alpha}^*\|_0 = |\mathcal{E}| = L$.

Proof: See the Appendix. ■

By exploiting the properties presented in Lemma 2, we derive an asymptotic upper bound on $\mathrm{WADD}(\tau_W(\boldsymbol{\alpha}^*, b))$. In particular, we have the following theorem:

Theorem 4: Let $\boldsymbol{\alpha}^*$ be defined as in (20). Assume that

$$\max_{\boldsymbol{E}\in\mathcal{E}} \mathbb{E}_{p_E}\left[\left(\log\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)^2\right] < \infty \tag{26}$$

We then have that as $b \to \infty$

$$\mathrm{WADD}\big(\tau_W(\boldsymbol{\alpha}^*, b)\big) \le \frac{b}{I_{\boldsymbol{\alpha}^*}}(1 + o(1)). \tag{27}$$

Proof: The proof is based on Lemma 2 and the analysis in [9] and is provided in the Appendix. ■

C. Asymptotic Optimality of M-CUSUM Test

By combining Theorems 3 with 4 we can establish the asymptotic optimality of the M-CUSUM test for weight choice $\boldsymbol{\lambda} = \boldsymbol{\alpha}^*$.

Theorem 5: Let $\boldsymbol{\alpha}^*$ be defined as in (20), and assume that

$$\max_{\boldsymbol{E}\in\mathcal{E}} \mathbb{E}_{p_{\boldsymbol{E}}}\left[\left(\log\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)^2\right] < \infty. \tag{28}$$

We then have:

i) For any $\gamma > 1, \mathbb{E}_\infty[\tau_W(\boldsymbol{\alpha}^*, \log\gamma)] \ge \gamma$.

ii) $\inf_{\tau\in\mathcal{C}_\gamma} \mathrm{WADD}(\tau) \sim \mathrm{WADD}(\tau_W(\boldsymbol{\alpha}^*, \log\gamma)) \sim \frac{\log\gamma}{I_{\boldsymbol{\alpha}^*}}$ as $\gamma \to \infty$.

Proof: i) follows from the MTFA analysis of the CUSUM test [6], [9], and ii) follows from Theorems 3 and 4. ■

Essentially, Theorem 5 implies that, for the case of heterogeneous sensors, there exists a choice of $\boldsymbol{\alpha}$ such that the M-CUSUM test that solves the QCD problem of (10)–(13) for said $\boldsymbol{\alpha}$ exactly is also asymptotically optimal with respect to (3)–(8). This $\boldsymbol{\alpha}$ is the one that minimizes the KL-divergence in (12).

The asymptotic optimality of the M-CUSUM test with weights $\boldsymbol{\alpha}^*$ can be intuitively explained through Lemma 2. In particular, since a larger γ implies a larger threshold, if we consider the logarithm of the M-CUSUM test statistic in (14), the expectation of the added log-likelihood ratio (which is usually referred to as the "drift" of the statistic) dominates the asymptotic performance of the M-CUSUM test. For a general choice of $\boldsymbol{\lambda}$ this drift is not generally equal for the different anomaly placements $\boldsymbol{E} \in \mathcal{E}$. Therefore, the worst-path delay will be dominated by the smallest resulting drift among anomaly placements. However, by Lemma 2 we know that choosing $\boldsymbol{\lambda} = \boldsymbol{\alpha}^*$ implies that the drift of the statistic is equal among a specific subset of anomaly placements. Furthermore, as we see in Lemma 2, all other placements of anomalous nodes lead to a larger drift, hence, do not play a role asymptotically due to the worst-path aspect of the delay. As a result, we have that the delay of our proposed test will match the universally best delay asymptotically.

Remark 3: Note that the first-order asymptotic optimality result presented in Theorem 5 also holds if we use a worst-path version of Pollak's detection delay [7]. In particular, for stopping time τ define the detection delay

$$\mathrm{CADD}(\tau) \triangleq \sup_{\boldsymbol{S}} \sup_{\nu\ge 0} \mathbb{E}_\nu^{\boldsymbol{S}}[\tau - \nu | \tau > \nu]. \tag{29}$$

By deriving a lower bound similar to the one in Theorem 3, and since WADD is always larger than CADD, we can easily establish the first-order asymptotic optimality of the M-CUSUM test under Pollak's criterion, i.e., Theorem 5 also holds when WADD is replaced by CADD.

VI. Numerical Results

In this section, we present numerical results for the studied moving anomaly QCD problem for the case of a single anomalous node ($m = 1$) and different network sizes L. We present results for both homogeneous and heterogeneous sensor networks.

Define the Gaussian distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu, \sigma^2)$. For the case of a homogeneous network, we assume that $g = \mathcal{N}(0, 1)$ and $f = \mathcal{N}(1, 1)$. For homogeneous networks, we can introduce two additional tests that can be used for comparison: a heuristic test; and an oracle-type test. In particular, note that for all $\boldsymbol{S}$ we have that

$$\mathbb{E}_\infty\left[\sum_{\ell=1}^{L}\log\frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)D(f\|g)\right] = -mD(f\|g) < 0$$

$$\mathbb{E}_0^{\boldsymbol{S}}\left[\sum_{\ell=1}^{L}\log\frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)D(f\|g)\right] = mD(f\|g) > 0.$$

This suggests that the following Naive-CUSUM (N-CUSUM) test may be a candidate test for detecting the distribution change described in (3).

$$W_N[k] \triangleq \left(W_N[k-1] + \sum_{\ell=1}^{L}\log\frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)D(f\|g)\right)^+$$

with $W_N[0] \triangleq 0$ and corresponding stopping time

$$\tau_N = \inf\{k \ge 1 : W_N[k] \ge b\}.$$

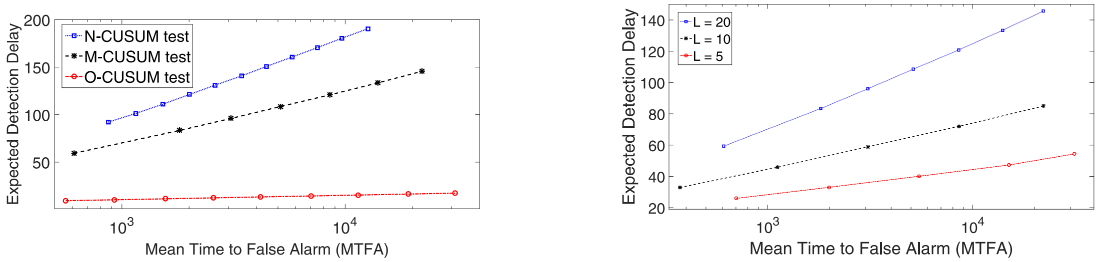
Although the N-CUSUM test can be employed to detect the anomaly due to the statistic $W_N[k]$ having a negative drift before and a positive drift after the change, it may be far from optimal for the QCD problem of interest defined in (8).

We also compare our proposed procedure to an Oracle-CUSUM (O-CUSUM) test, which is a CUSUM test that uses complete knowledge of $\boldsymbol{S}$. That is, to define this test we assume that at time k we do not know whether a change has occured, but we know which set of sensors would be affected if an anomaly had already emerged in the network. In particular, consider the statistic calculated by using the following recursion:

$$W_O[k] = \left(W_O[k-1] + \log\left(\prod_{\ell\in\boldsymbol{S}[k]}\frac{f(X_\ell[k])}{g(X_\ell[k])}\right)\right)^+ \tag{30}$$

with $W_O[0] \triangleq 0$ and with corresponding stopping time

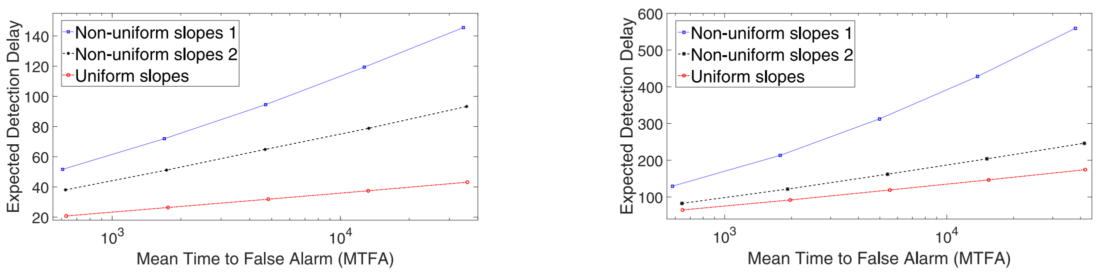
$$\tau_O = \inf\{k \ge 1 : W_O[k] \ge b\}. \tag{31}$$



(a) WADD versus MTFA for $L = 20$, $m = 1$.

(b) WADD versus MTFA for the M-CUSUM when $m = 1$ and for different L values.

Fig. 1. WADD versus MTFA for homogeneous sensor network.



(a) WADD versus MTFA for $L = 10$, $m = 1$.

(b) WADD versus MTFA for $L = 20$, $m = 1$.

Fig. 2. WADD versus MTFA for heterogeneous sensor network.

Since this O-CUSUM test uses the knowledge of the location of the anomalous nodes, it is expected to perform better than our proposed test. However, such a test is not implementable since in practice such location information will not be available to the decision maker.

In Fig. 1(a) we compare the M-CUSUM test, with the N-CUSUM test and the O-CUSUM test for network size $L = 20$. Note that due to the symmetry of the M-CUSUM and the N-CUSUM test statistics, WADD is equal to the delay for any arbitrary path of the anomaly. By inspecting Fig. 1(a) we note that the M-CUSUM test outperforms the heuristic N-CUSUM test, which is expected since the M-CUSUM test is optimal with respect to (8). In addition, we note that the O-CUSUM test performs better than the other detection schemes, which is to be expected since it exploits complete knowledge of $\boldsymbol{S}$. In Fig. 1(b), we evaluate the performance of our proposed M-CUSUM test for different values of L. We note that as L increases our proposed test performs worse, which is expected since the algorithm is affected by more "noise" from non-anomalous nodes for larger network sizes. It should be noted that the graph points presented thus far, as well as in the remainder of the paper, were generated by varying the threshold used to account for different values of MTFA and delay. In practice, the test threshold can be chosen to satisfy the MTFA constraint by setting $b = \log \gamma$, as highlighted in the analysis of the M-CUSUM test.

For the case of a heterogeneous sensor network, we compare three versions of the test introduced in (14)–(16): the first version ("Uniform slopes" in Fig. 2) uses the optimal weights $\boldsymbol{\alpha}^*$ to achieve a uniform average statistic drift among anomaly placements (see Lemma 2); the second and third versions ("Non-uniform slopes 1" and "Non-uniform slopes 2" in Fig. 2) use arbitrary choices of weights that only guarantee that the expected drift of the statistic is positive for any placement of the anomaly. The optimal weights are found by using gradient descent with the derivatives calculated as in (59). Note that each derivative is equal to a difference of two expected values, which we calculate using Monte Carlo methods. Note that this method of calculating $\boldsymbol{\alpha}^*$ can be computationally expensive, especially for specific values of m and large network sizes L. Alternative techniques can be derived by using approximations of the KL divergence between two Gaussian mixture models (see, e.g., [33]). In addition, note that $\boldsymbol{\alpha}^*$ is calculated offline, hence does not affect the calculation of the test statistic.

It should be noted that the WADD in the case of heterogeneous sensor networks is calculated approximately, since the worst path of the anomaly cannot be specified analytically. However, as the MTFA becomes large, the WADD can be approximated by placing the anomalies at only the nodes (in this case node since $m = 1$) that correspond to the worst post-change drift for the test statistic. For the optimal weight choice, the placement of the anomaly does not affect the delay for large MTFA, since the drift does not depend on the trajectory of the anomaly.

We consider the cases of $L = 10$ and $L = 20$. For the case of $L = 10$, we assume that $g_\ell = \mathcal{N}(0, 1)$ for all $\ell \in [L]$, and that $f_\ell = \mathcal{N}(\mu_\ell, 1)$ 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 post-change means. The results can be seen in Fig. 2(a). The M-CUSUM test statistic using optimal weights is then characterized by a uniform statistic drift, approximately equal to 0.178. For the case of "Non-uniform slopes 1" the worst drift corresponds to placing the anomaly at sensor 2, corresponding to an approximate slope of 0.029, and for the case of "Non-uniform slopes 2" at sensor 5, with an approximate slope of 0.065. We see that he mixture-CUSUM test using the optimal weights $\boldsymbol{\alpha}^*$ outperforms the other two implementations. Similar results can be produced by considering the case of $L = 20$. For that case, we assume that $g_\ell = \mathcal{N}(0, 1)$ for all $\ell \in [L]$, $f_\ell = \mathcal{N}(0.8, 1)$ for all $1 \leq \ell \leq 5$, $f_\ell = \mathcal{N}(1, 1)$ for all $6 \leq \ell \leq 15$, and $f_\ell = \mathcal{N}(1.2, 1)$ for all $16 \leq \ell \leq 20$. The results can be seen in Fig. 2(b), where we note that the optimal weights test outperforms the tests that use arbitrarily chosen weights. The resulting homogeneous statistic drift is then approximately equal to 0.036. Furthermore, for the case of "Non-uniform slopes 1" the worst drift corresponds to placing the anomaly at any sensor $\ell \in [5]$, corresponding to an approximate slope of 0.003, and for the case of "Non-uniform slopes 2" at any sensor $\ell \in \{16, 17, 18, 19, 20\}$, with an approximate slope equal to 0.023. Finally, it should be noted that in this case we have chosen "Non-uniform slopes 1" to correspond to the case of uniform weights. As a result, the gap between the blue and red lines in Fig. 2(b) captures the loss we suffer if we make the assumption that the sensors of the network are homogeneous.

VII. Conclusion

We studied the problem of moving anomaly detection, where an anomaly emerges in a sensor network affecting different nodes at each time instant after its appearance. We posed the problem in a minimax QCD setting, where the trajectory of the anomaly is assumed to be *unknown* but *deterministic*. To this end, we introduced a modified version of Lorden's [6] detection delay metric that evaluates candidate detection schemes according to the worst performance with respect to the path of the anomaly. We proposed a CUSUM-type test that is an exact solution to the moving anomaly QCD problem for the case of a homogeneous network, and is also first-order asymptotically optimal when applied to a heterogeneous network.

Future work in this area includes studying the case of a moving anomaly of size varying with time (for current progress on this problem see [34], modifying the proposed procedures to provide robustness with respect to limited knowledge of data-generating distributions, as well as, studying the case of moving anomaly detection under the presence of adversarial attacks.

Appendix

Proof of Theorem 1: For any stopping time τ adapted to $\mathscr{F}$ and $N > 0$ define the truncated version of τ by $\tau^{(N)} \triangleq \min\{\tau, N\}$. Fix $\boldsymbol{\alpha} \in \mathcal{A}$. Due to the presence of the sup and ess sup in (6), we have that for any path $\boldsymbol{S}$, $\nu \geq 0$, $\mathscr{F}$-adapted stopping time τ and $N > 0$

$$\begin{aligned}\mathrm{WADD}\Big(\tau^{(N)}\Big) &\geq \mathbb{E}_\nu^{\boldsymbol{S}}\Big[\tau^{(N)} - \nu | \tau^{(N)} > \nu, \mathscr{F}_\nu\Big] \\ &= \mathbb{E}_\nu^{\boldsymbol{S}}\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} > \nu, \mathscr{F}_\nu\right] \\ &\overset{(a)}{=} \mathbb{E}_\infty\left[\sum_{j=\nu}^{\infty} \Gamma_{\boldsymbol{S}}(j,\nu) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} > \nu, \mathscr{F}_\nu\right]\end{aligned} \tag{32}$$

where (a) follows by changing the measure to $\mathbb{P}_\infty(\cdot)$. By multiplying both sides of the inequality (32) with $\mathbb{1}_{\{\tau^{(N)} > \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu])^+$ and taking the expected value under $\mathbb{E}_\infty[\,\cdot\,]$ we then have that

$$\begin{aligned}&\mathbb{E}_\infty\Big[\mathbb{1}_{\{\tau^{(N)} > \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \mathrm{WADD}\Big(\tau^{(N)}\Big)\Big] \\ &\quad = \mathbb{E}_\infty\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} > \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \Gamma_{\boldsymbol{S}}(j,\nu) \mathbb{1}_{\{\tau^{(N)} > j\}}\right]\end{aligned}$$

By summing both sides of this equation over ν from $\nu = 0$ to $\nu = N$, and due to the linearity of expectation and the fact that $\tau^{(N)} \leq N$ we have that

$$\begin{aligned}&\mathbb{E}_\infty\left[\sum_{\nu=0}^{\tau^{(N)}-1}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \mathrm{WADD}\Big(\tau^{(N)}\Big)\right] \\ &\quad \geq \mathbb{E}_\infty\left[\sum_{\nu=0}^{\tau^{(N)}-1}\sum_{j=\nu}^{\tau^{(N)}-1}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \Gamma_{\boldsymbol{S}}(j,\nu)\right] \\ &\quad = \mathbb{E}_\infty\left[\sum_{j=0}^{\tau^{(N)}-1}\sum_{\nu=0}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \Gamma_{\boldsymbol{S}}(j,\nu)\right].\end{aligned}$$

By taking the sup with respect to $\boldsymbol{S}$, we obtain

$$\begin{aligned}&\mathrm{WADD}\Big(\tau^{(N)}\Big) \\ &\quad \geq \frac{\sup_{\boldsymbol{S}[1,N-1]} \mathbb{E}_\infty\Big[\sum_{j=0}^{\tau^{(N)}-1}\sum_{\nu=0}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \Gamma_{\boldsymbol{S}}(j,\nu)\Big]}{\mathbb{E}_\infty\Big[\sum_{\nu=0}^{\tau^{(N)}-1}(1 - W_{\boldsymbol{\alpha}}[\nu])^+\Big]}\end{aligned} \tag{33}$$

To proceed, we further bound the numerator in (33). For $1 \leq n < N$, define the following function

$$\begin{aligned}&\Phi_{n,N-1}(\boldsymbol{S}[1, n-1], \boldsymbol{S}[n+1, N-1]) \\ &\quad \triangleq \sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty\left[\sum_{j=0}^{N-1}\sum_{\nu=0}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \Gamma_{\boldsymbol{S}}(j,\nu) \mathbb{1}_{\{\tau^{(N)} > j\}}\right].\end{aligned} \tag{34}$$

Then, by first taking the sup over $\boldsymbol{S}[n]$ we have that

$$\begin{aligned}&\sup_{\boldsymbol{S}[1,N-1]} \mathbb{E}_\infty\left[\sum_{j=0}^{N-1}\sum_{\nu=0}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu])^+ \Gamma_{\boldsymbol{S}}(j,\nu) \mathbb{1}_{\{\tau^{(N)} > j\}}\right] \\ &\quad = \sup_{\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]} \Phi_{n,N-1}(\boldsymbol{S}[1, n-1], \boldsymbol{S}[n+1, N-1]).\end{aligned} \tag{35}$$

For $0 \leq j < N$ and $0 \leq n < N$ define

$$\begin{aligned}A_{j,n} \triangleq &\left(\sum_{\nu=0}^{n-1}(1 - W_{\boldsymbol{\alpha}}[\nu])^+\right. \\ &\left.\left(\prod_{\substack{i=\nu+1 \\ i \neq n}}^{j} \Gamma_{\boldsymbol{S}}(i, i-1)\right) \mathbb{1}_{\{\tau^{(N)} > j\}}\right) \mathbb{1}_{\{j \geq n\}}\end{aligned} \tag{36}$$

and

$$B_{j,n} \triangleq \left(\sum_{\nu=0}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\Gamma_{\boldsymbol{S}}(j,\nu)\mathbb{1}_{\{\tau^{(N)}>j\}}\right)\mathbb{1}_{\{j<n\}} + \left(\sum_{\nu=n}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\Gamma_{\boldsymbol{S}}(j,\nu)\mathbb{1}_{\{\tau^{(N)}>j\}}\right)\mathbb{1}_{\{j\geq n\}}. \quad (37)$$

It can then be shown that for any $0 \leq n < N$

$$\sum_{\nu=0}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\Gamma_{\boldsymbol{S}}(j,\nu)\mathbb{1}_{\{\tau^{(N)}>j\}} = \Gamma_{\boldsymbol{S}}(n,n-1)A_{j,n}+B_{j,n}. \quad (38)$$

Then from (34), (38) we have that

$$\Phi_{n,N-1}(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]) = \sup_{\boldsymbol{S}[n]}\mathbb{E}_{\infty}\left[\Gamma_{\boldsymbol{S}}(n,n-1)\sum_{j=0}^{N-1}A_{j,n}+\sum_{j=0}^{N-1}B_{j,n}\right]. \quad (39)$$

Note that since $A_{j,n}$ and $B_{j,n}$ are independent of $\boldsymbol{S}[n]$ under $\mathbb{P}_{\infty}(\cdot)$, we have that for all

$$\boldsymbol{E}\in\mathcal{E} = \sup_{\boldsymbol{S}[n]}\mathbb{E}_{\infty}\left[\left(\prod_{\ell\in\boldsymbol{S}[n]}\frac{f_{\ell}(X_{\ell}[n])}{g_{\ell}(X_{\ell}[n])}\right)\sum_{j=0}^{N-1}A_{j,n}+\sum_{j=0}^{N-1}B_{j,n}\right] \geq \mathbb{E}_{\infty}\left[\left(\prod_{\ell\in\boldsymbol{E}}\frac{f_{\ell}(X_{\ell}[n])}{g_{\ell}(X_{\ell}[n])}\right)\sum_{j=0}^{N-1}A_{j,n}+\sum_{j=0}^{N-1}B_{j,n}\right] \quad (40)$$

which together with (39) implies that

$$\Phi_{n,N-1}(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]) \geq \mathbb{E}_{\infty}\left[\left(\prod_{\ell\in\boldsymbol{E}}\frac{f_{\ell}(X_{\ell}[n])}{g_{\ell}(X_{\ell}[n])}\right)\sum_{j=0}^{N-1}A_{j,n}+\sum_{j=0}^{N-1}B_{j,n}\right]. \quad (41)$$

By averaging both sides of (41) with respect to $\boldsymbol{\alpha}$ we then have that

$$\begin{aligned}\Phi_{n,N-1}&(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1])\\ &\geq \mathbb{E}_{\infty}\left[\left(\sum_{\boldsymbol{E}\in\mathcal{E}}\alpha_{\boldsymbol{E}}\left(\prod_{\ell\in\boldsymbol{E}}\frac{f_{\ell}(X_{\ell}[n])}{g_{\ell}(X_{\ell}[n])}\right)\right)\sum_{j=0}^{N-1}A_{j,n}+\sum_{j=0}^{N-1}B_{j,n}\right]\\ &= \mathbb{E}_{\infty}\left[\sum_{j=0}^{N-1}\sum_{\nu=0}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\mathcal{L}_{\boldsymbol{\alpha}}(n,n-1)\left(\prod_{\substack{i=\nu+1\\ i\neq n}}^{j-1}\Gamma_{\boldsymbol{S}}(i,i-1)\right)\mathbb{1}_{\{\tau^{(N)}>j\}}\right]. \quad (42)\end{aligned}$$

By unfolding (35) in the same fashion with respect to all $0 \leq n < N$, it can be easily shown that

$$\sup_{\boldsymbol{S}[1,N-1]}\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\sum_{\nu=0}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\Gamma_{\boldsymbol{S}}(j,\nu)\right] \geq \mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\sum_{\nu=0}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\mathcal{L}_{\boldsymbol{\alpha}}(j,\nu)\right] \quad (43)$$

which in turn together with (33) implies that

$$\mathrm{WADD}\left(\tau^{(N)}\right) \geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\sum_{\nu=0}^{j}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\mathcal{L}_{\boldsymbol{\alpha}}(j,\nu)\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau^{(N)}-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}.$$

From [8, Lemma 1] we have that

$$\sum_{\nu=0}^{j-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\mathcal{L}_{\boldsymbol{\alpha}}(j,\nu) = W_{\boldsymbol{\alpha}}[j]$$

which together with the previous equation implies that

$$\mathrm{WADD}\left(\tau^{(N)}\right) \geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\left(W_{\boldsymbol{\alpha}}[j]+(1-W_{\boldsymbol{\alpha}}[j])^{+}\right)\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau^{(N)}-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]} = \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\max\{W_{\boldsymbol{\alpha}}[j],1\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau^{(N)}-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}. \quad (44)$$

Consider b chosen such that $\mathbb{E}_{\infty}[\tau_W(\boldsymbol{\alpha},b)] = \gamma$. Let $b' \geq b$ such that $b' > 0$. Then

$$\begin{aligned}\mathrm{WADD}(\tau) &\geq \mathrm{WADD}\left(\tau^{(N)}\right)\\ &\geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\max\{W_{\boldsymbol{\alpha}}[j],1\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau^{(N)}-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}\\ &\geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau^{(N)}-1}\min\left\{\max\{W_{\boldsymbol{\alpha}}[j],1\},e^{b'}\right\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau^{(N)}-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}.\end{aligned}$$

Since $\mathbb{E}_{\infty}[\tau] < \infty$, taking the limit as $n \to \infty$ we have that

$$\mathrm{WADD}(\tau) \geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau-1}\min\left\{\max\{W_{\boldsymbol{\alpha}}[j],1\},e^{b'}\right\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}. \quad (45)$$

Since (45) holds for arbitrary τ adapted to $\mathscr{F}$, we have that for any $\gamma > 1$

$$\inf_{\tau\in C_{\gamma}}\mathrm{WADD}(\tau) \geq \frac{\inf_{\tau\in C_{\gamma}}\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau-1}\min\left\{\max\{W_{\boldsymbol{\alpha}}[j],1\},e^{b'}\right\}\right]}{\sup_{\tau\in C_{\gamma}}\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}. \quad (46)$$

Note that the function $\phi(x) = (1-x)^{+}$ in continuous and non-increasing with $\phi(0) = 1$. Furthermore, note that the function $\psi(x) = -\min\{\max\{x,1\},e^{b'}\}$ is continuous and non-increasing in x with $\psi(0) = -\min\{1,e^{b'}\}$. As a result, from [8, Th. 1] we also have that

$$\inf_{\tau\in C_{\gamma}}\mathrm{WADD}(\tau) \geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau_W(\boldsymbol{\alpha},b)-1}\min\left\{\max\{W_{\boldsymbol{\alpha}}[j],1\},e^{b'}\right\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau_W(\boldsymbol{\alpha},b)-1}(1-W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}$$

$$\stackrel{(d)}{=} \frac{\mathbb{E}_\infty\left[\sum_{j=0}^{\tau_W(\boldsymbol{\alpha},b)-1} \max\{W_{\boldsymbol{\alpha}}[j], 1\}\right]}{\mathbb{E}_\infty\left[\sum_{\nu=0}^{\tau_W(\boldsymbol{\alpha},b)-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^+\right]} \tag{47}$$

where (d) is implied since $W_{\boldsymbol{\alpha}}[j] < e^b \leq e^{b'}$ for $0 \leq j < \tau_W(\boldsymbol{\alpha}, b)$ and since $b' > 0$. Furthermore, note that from the optimality of the CUSUM test for the classic QCD problem [8] we have that

$$\frac{\mathbb{E}_\infty\left[\sum_{j=0}^{\tau_W(\boldsymbol{\alpha},b)-1} \max\{W_{\boldsymbol{\alpha}}[j], 1\}\right]}{\mathbb{E}_\infty\left[\sum_{\nu=0}^{\tau_W(\boldsymbol{\alpha},b)-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^+\right]} = \overline{\text{WADD}}(\tau_W(\boldsymbol{\alpha}, b)). \tag{48}$$

As a result, from (47) and (48) and since

$$\text{WADD}(\tau_W(\boldsymbol{\alpha}, b)) \geq \inf_{\tau \in C_\gamma} \text{WADD}(\tau) \tag{49}$$

the theorem is established.

Proof of Lemma 1: Fix $\boldsymbol{\alpha} \in \mathcal{A}$, $b > 0$ and $N > 0$. For purposes of presentation of this proof, we denote the stopping $\tau_W(\boldsymbol{\lambda}_U, b)$ with uniform weights and threshold b by simply τ_W and $W_{\boldsymbol{\lambda}_U}[k]$, $\mathcal{L}_{\boldsymbol{\lambda}_U}(\cdot,\cdot)$ by $W[k]$ and $\mathcal{L}(\cdot,\cdot)$ respectively. Define the truncated stopping time $\tau_W^{(N)} = \min\{\tau_W, N\}$. Note that by employing a change of measure similar to the one in (32) we have that for any $\nu \geq 0$ and any $\boldsymbol{S}$

$$\begin{aligned} V_\nu &\triangleq \mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau_W^{(N)} - \nu \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right] \\ &= 1 + \mathbb{E}_\infty\left[\sum_{j=\nu+1}^{N-1} \Gamma_{\boldsymbol{S}}(j,\nu) \mathbb{1}_{\left\{\tau_W^{(N)} > j\right\}} \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right] \\ &= 1 + \mathbb{E}_\infty\left[\sum_{j=\nu+1}^{N-1} \Gamma_{\boldsymbol{S}}(j,\nu) \left(\prod_{i=\nu+1}^{j} \mathbb{1}_{\{W[i] < e^b\}}\right) \right. \\ &\qquad\qquad \left. \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right]. \end{aligned} \tag{50}$$

Furthermore, by using induction it can be seen that for any $0 \leq \nu \leq N-1$

$$V_\nu = 1 + \mathbb{E}_\infty\left[\Gamma_{\boldsymbol{S}}(\nu+1,\nu) \mathbb{1}_{\{W[\nu+1] < e^b\}} V_{\nu+1} \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right]$$

with $V_\nu = 1$ for all $\nu \geq N-1$.

By further analyzing V_ν it can be shown that V_ν is independent of $\boldsymbol{S}$ for all $\nu \geq 0$ and that it is a function of $\mathscr{F}_\nu$ only through $W[\nu]$. This implies that for $\nu \geq 0$ and $\boldsymbol{E} \in \mathcal{E}$

$$\begin{aligned} V_\nu = 1 + \mathbb{E}_\infty&\left[\left(\prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[\nu+1])}{g(X_\ell[\nu+1])}\right) \mathbb{1}_{\{W[\nu+1] < e^b\}} V_{\nu+1} \right. \\ &\qquad \left. \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right]. \end{aligned} \tag{51}$$

As a result, by averaging over $\boldsymbol{E}$ with respect to $\boldsymbol{\alpha}$ we have that

$$V_\nu = 1 + \mathbb{E}_\infty\left[\mathcal{L}(\nu+1,\nu) \mathbb{1}_{\{W[\nu+1] < e^b\}} V_{\nu+1} \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right].$$

By unfolding this recursion, it can be easily seen that for any $\nu \geq 0$ and any $\boldsymbol{S}$

$$\mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau_W^{(N)} - \nu | \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right] = \overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}\left[\tau_W^{(N)} - \nu | \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right]. \tag{52}$$

From the Monotone Convergence Theorem, since $\tau_W^{(N)} - \nu$ and $\mathbb{1}_{\{\tau_W^{(N)} - \nu\}}$ are non-decreasing with N, we have that for all $\boldsymbol{S}$

$$\begin{aligned} &\lim_{N\to\infty} \mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau_W^{(N)} - \nu \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right] \\ &\quad = \mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau_W - \nu | \tau_W > \nu, \mathscr{F}_\nu\right]. \end{aligned} \tag{53}$$

Similarly, it can be shown that

$$\begin{aligned} &\lim_{N\to\infty} \overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}\left[\tau_W^{(N)} - \nu \middle| \tau_W^{(N)} > \nu, \mathscr{F}_\nu\right] \\ &\quad = \overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}\left[\tau_W - \nu | \tau_W > \nu, \mathscr{F}_\nu\right]. \end{aligned} \tag{54}$$

As a result, by taking the limit on both sides of (52) and using (53) and (54) we have that for all $\nu \geq 0$, $\boldsymbol{S}$

$$\mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau_W - \nu | \tau_W > \nu, \mathscr{F}_\nu\right] = \overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}\left[\tau_W - \nu | \tau_W > \nu, \mathscr{F}_\nu\right] \tag{55}$$

which in turn implies

$$\text{WADD}(\tau_W) = \overline{\text{WADD}}_{\boldsymbol{\alpha}}(\tau_W). \tag{56}$$

Proof of Lemma 2: Define $\boldsymbol{\beta} = [\beta_{\boldsymbol{E}_1}, \ldots, \beta_{\boldsymbol{E}_{|\mathcal{E}|-1}}]^\top$ where $\alpha_{\boldsymbol{E}_j} \triangleq \beta_{\boldsymbol{E}_j}$ for $j \in [|\mathcal{E}| - 1]$. The constrained optimization of $\boldsymbol{I}_{\boldsymbol{\alpha}}$ can then be equivalently replaced by

$$\begin{aligned} \inf_{\boldsymbol{\beta}} \quad & q(\boldsymbol{\beta}) \\ \text{s.t.} \quad & \beta_{\boldsymbol{E}_j} \geq 0, \ \forall\, j \in [|\mathcal{E}| - 1] \\ & \sum_{j=1}^{|\mathcal{E}|-1} \beta_{\boldsymbol{E}_j} \leq 1, \end{aligned} \tag{57}$$

where

$$\begin{aligned} q(\boldsymbol{\beta}) \triangleq \int_{\mathbb{R}^L} &\left((1 - \|\boldsymbol{\beta}\|_1) p_{\boldsymbol{E}_{|\mathcal{E}|}}(\boldsymbol{x}) + \sum_{j=1}^{|\mathcal{E}|-1} \beta_{\boldsymbol{E}_j} p_{\boldsymbol{E}_j}(\boldsymbol{x})\right) \\ &\log\left(\frac{\left((1 - \|\boldsymbol{\beta}\|_1) p_{\boldsymbol{E}_{|\mathcal{E}|}}(\boldsymbol{x}) + \sum_{j=1}^{|\mathcal{E}|-1} \beta_{\boldsymbol{E}_j} p_{\boldsymbol{E}_j}(\boldsymbol{x})\right)}{g(\boldsymbol{x})}\right) d\boldsymbol{x}. \end{aligned} \tag{58}$$

Denote by $\boldsymbol{\beta}^*$ the solution to (57). Then, the derivative at $\boldsymbol{\beta}^*$ is given by

$$\begin{aligned} \left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\boldsymbol{\beta}^*} &= \mathbb{E}_{p_{\boldsymbol{E}_i}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \\ &\quad - \mathbb{E}_{p_{\boldsymbol{E}_{|\mathcal{E}|}}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right]. \end{aligned} \tag{59}$$

Without loss of generality we have that either $\boldsymbol{\beta}^* = [\beta_{\boldsymbol{E}_1}^*, \ldots, \beta_{\boldsymbol{E}_\eta}^*, \ldots, 0]^\top$ with $\eta \in [|\mathcal{E}| - 1]$ and $\beta_{\boldsymbol{E}_j}^* > 0$ for all $j \in [\eta]$ (boundary or interior point), or $\boldsymbol{\beta}^* = [0, \ldots, 0]^\top$ (corner point).

Assume that $\boldsymbol{\beta}^*$ is a corner point. Denote by $D(f\|g)$ denote the KL-divergence between two pdfs $f(\cdot)$ and $g(\cdot)$. In this case we have that for all $i \in [|\mathcal{E}|-1]$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\boldsymbol{\beta}^*} = \sum_{\ell\in \boldsymbol{E}_{|\mathcal{E}|}} \big(D(f_\ell\|g_\ell)\mathbb{1}_{\{\ell\in\boldsymbol{E}_i\}} - D(g_\ell\|f_\ell)\mathbb{1}_{\{\ell\notin\boldsymbol{E}_i\}}\big) - \sum_{\ell\in \boldsymbol{E}_{|\mathcal{E}|}} D(f_i\|g_i) < 0, \tag{60}$$

which is a contradiction since

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\boldsymbol{\beta}^*} \geq 0 \tag{61}$$

must hold for all $i \in [|\mathcal{E}|-1]$ due to the fact that $\boldsymbol{\beta}^*$ is a minimum. As a result, $\boldsymbol{\beta}^*$ is not a corner point. In this case, for all $i \in [\eta]$ we have that

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\boldsymbol{\beta}^*} = 0, \tag{62}$$

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

$$\mathbb{E}_{p_{\boldsymbol{E}_i}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = \mathbb{E}_{p_{\boldsymbol{E}_{|\mathcal{E}|}}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \triangleq J. \tag{63}$$

Furthermore, we have that since $\alpha^*_{\boldsymbol{E}_j} = 0$ for $\eta < j < |\mathcal{E}|$

$$\begin{aligned} J &= \left(\sum_{j=1}^{\eta}\beta^*_{\boldsymbol{E}_j} + \left(1-\sum_{j=1}^{\eta}\beta^*_{\boldsymbol{E}_j}\right)\right)J \\ &= \left(\sum_{j=1}^{\eta}\alpha^*_{\boldsymbol{E}_j} + \alpha^*_{\boldsymbol{E}_{|\mathcal{E}|}}\right)J \\ &= \sum_{j=1}^{|\mathcal{E}|}\alpha^*_{\boldsymbol{E}_j}\mathbb{E}_{p_{\boldsymbol{E}_j}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \\ &= \mathbb{E}_{\overline{p}_{\boldsymbol{\alpha}^*}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \\ &= I_{\boldsymbol{\alpha}^*} > 0. \end{aligned} \tag{64}$$

In addition, we have that for $\eta < i < |\mathcal{E}|$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\boldsymbol{\beta}^*} > 0. \tag{65}$$

This implies that for all $i \in [\eta]\cup\{|\mathcal{E}|\}$ and $\eta < j < |\mathcal{E}|$

$$\mathbb{E}_{p_{\boldsymbol{E}_j}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] > \mathbb{E}_{p_{\boldsymbol{E}_i}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] = I_{\boldsymbol{\alpha}^*}. \tag{66}$$

ii) For the case of $m = 1$, without loss of generality assume that for all $1 \leq j \leq |\mathcal{E}| = L$, we have that $E_j = j$. For $\eta < i < L$, we then have that

$$\begin{aligned} \mathbb{E}_{p_{E_i}}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] &= \mathbb{E}_{p_i}\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] \\ &= \mathbb{E}_g\left[\log\left(\sum_{j=1}^{\eta}\alpha^*_j\frac{f_j(X_j)}{g_j(X_j)} + \alpha^*_L\frac{f_L(X_L)}{g_L(X_L)}\right)\right] \\ &= \mathbb{E}_g\left[\log\left(\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right)\right] < 0. \end{aligned} \tag{67}$$

We then have that from (59), (63), (64) and (67)

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{E_i}}\right|_{\boldsymbol{\beta}^*} < 0 \tag{68}$$

for all $\eta < i < L$, which leads to a contradiction, since (68) cannot hold at the minimum.

Proof of Theorem 4: Our upper bound analysis is based on the proof technique in [9]. Due to the structure of the test we have that for any $b > 0$

$$\mathrm{WADD}\big(\tau_W(\boldsymbol{\alpha}^*, b)\big) = \sup_{\boldsymbol{S}}\mathbb{E}_0^{\boldsymbol{S}}\big[\tau_W(\boldsymbol{\alpha}^*, b)\big]. \tag{69}$$

Let $0 < \epsilon < I_{\boldsymbol{\alpha}^*}$ and $n_b = \frac{b}{I_{\boldsymbol{\alpha}^*}-\epsilon}$. We then have that

$$\begin{aligned} \sup_{\boldsymbol{S}}\mathbb{E}_0^{\boldsymbol{S}}\left[\frac{\tau_W(\boldsymbol{\alpha}^*, b)}{n_b}\right] &\overset{(a)}{=} \sup_{\boldsymbol{S}}\int_0^{\infty}\mathbb{P}_0^{\boldsymbol{S}}\left(\frac{\tau_W(\boldsymbol{\alpha}^*, b)}{n_b} > x\right)dx \\ &\overset{(b)}{\leq} 1 + \sup_{\boldsymbol{S}}\lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi} \\ &\quad\times \mathbb{P}_0^{\boldsymbol{S}}\big(\tau_W(\boldsymbol{\alpha}^*, b) > \zeta n_b\big), \end{aligned} \tag{70}$$

where (a) follows from writing the expectation as an integral of the inverse cumulative density function for a positive random variable and (b) from the sum-integral inequality. Define the log-likelihood ratio at time j corresponding to (10) for $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ by

$$Z_{\boldsymbol{\alpha}^*}[j] \triangleq \log\frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}. \tag{71}$$

For any path $\boldsymbol{S} = \{\boldsymbol{S}[k]\}_{k=1}^{\infty}$, $\zeta \geq 1$, we then have that

$$\begin{aligned} \mathbb{P}_0^{\boldsymbol{S}}\big(\tau_W(\boldsymbol{\alpha}^*, b) > \zeta n_b\big) &= \mathbb{P}_0^{\boldsymbol{S}}\left(\max_{1\leq k\leq\zeta n_b} W_{\boldsymbol{\alpha}^*}[k] < e^b\right) \\ &\overset{(c)}{=} \mathbb{P}_0^{\boldsymbol{S}}\left(\max_{1\leq k\leq\zeta n_b}\max_{1\leq i\leq k}\mathcal{L}_{\boldsymbol{\alpha}^*}(k, i-1) < e^b\right) \\ &\overset{(d)}{\leq} \prod_{r=1}^{\zeta}\mathbb{P}_0^{\boldsymbol{S}}\left(\frac{\sum_{j=(r-1)n_b+1}^{rn_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right) \end{aligned} \tag{72}$$

where (c) follows from the definition of the M-CUSUM statistic (14) and after taking the logarithm at both sides of the inequality, and (d) follows by using the binning technique in [9] and by the independence of the observations over time. Note that for $b > 0$ we then have that from (72)

$$\begin{aligned} &\sup_{\boldsymbol{S}}\lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\mathbb{P}_0^{\boldsymbol{S}}\big(\tau_W(\boldsymbol{\alpha}^*, b) > \zeta n_b\big) \\ &\leq \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\sup_{\boldsymbol{S}}\mathbb{P}_0^{\boldsymbol{S}}\big(\tau_W(\boldsymbol{\alpha}^*, b) > \zeta n_b\big) \\ &\leq \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\prod_{r=1}^{\zeta}\left[\sup_{\boldsymbol{S}}\mathbb{P}_0^{\boldsymbol{S}}\left(\frac{\sum_{j=(r-1)n_b+1}^{rn_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right)\right] \\ &= \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\left[\sup_{\boldsymbol{S}}\mathbb{P}_0^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right)\right]^{\zeta}. \end{aligned} \tag{73}$$

For fixed $\boldsymbol{S}$, b define

$$I_{\boldsymbol{S},b} \triangleq \mathbb{E}_0^{\boldsymbol{S}}\left[\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b}\right] = \frac{\sum_{j=1}^{n_b}\mathbb{E}_{p_{\boldsymbol{S}[j]}}\big[Z_{\boldsymbol{\alpha}^*}[j]\big]}{n_b} \geq I^*_{\boldsymbol{\alpha}}, \tag{74}$$

where the inequality follows from Lemma 2. This in turn implies that for any $\boldsymbol{S}$ we have that

$$\mathbb{P}_0^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b} < I_{\boldsymbol{\alpha}^*} - \epsilon\right) \leq \mathbb{P}_0^{\boldsymbol{S}}\left(\left|\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b} - I_{\boldsymbol{S},b}\right| > \epsilon\right).$$

Define

$$\bar{\sigma}^2 \triangleq \max_{\boldsymbol{E}\in\mathcal{E}} \mathrm{Var}_{p_{\boldsymbol{E}}}\left[\log \frac{\overline{p}_{\boldsymbol{\alpha}^*}(\boldsymbol{X})}{g(\boldsymbol{X})}\right]. \tag{75}$$

From (26), we have that $\bar{\sigma}^2 < \infty$. Then, by Chebychev's inequality

$$\begin{aligned}\mathbb{P}_0^{\boldsymbol{S}}\left(\left|\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b} - I_{\boldsymbol{S},b}\right| > \epsilon\right) &\leq \mathrm{Var}_0^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}^*}[j]}{n_b}\right)\frac{1}{\epsilon^2} \\ &= \frac{1}{\epsilon^2 n_b^2}\sum_{j=1}^{n_b} \mathrm{Var}_{p_{\boldsymbol{S}[j]}}(Z_{\boldsymbol{\alpha}^*}[j]) \\ &\leq \frac{\sum_{j=1}^{n_b} \bar{\sigma}^2}{n_b^2\epsilon^2} = \frac{\bar{\sigma}^2}{n_b\epsilon^2}. \end{aligned} \tag{76}$$

By using (70), (73), (75) and (76) we then have that

$$\sup_{\boldsymbol{S}} \mathbb{E}_0^{\boldsymbol{S}}\left[\frac{\tau_W(\boldsymbol{\alpha}^*, b)}{n_b}\right] \leq 1 + \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\left[\frac{\bar{\sigma}^2}{n_b\epsilon^2}\right]^{\zeta}. \tag{77}$$

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

$$\sup_{\boldsymbol{S}} \mathbb{E}_0^{\boldsymbol{S}}\left[\frac{\tau_W(\boldsymbol{\alpha}^*, b)}{n_b}\right] \leq 1 + \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\delta^{\zeta} = \sum_{\zeta=0}^{\infty}\delta^{\zeta} = \frac{1}{1-\delta}. \tag{78}$$

Since (78) holds for all $\epsilon > 0$ and $\delta \to 0$ as $b \to \infty$, by the definition of n_b the theorem is established.

REFERENCES

[1] 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. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 98–102.

[2] 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)*, Los Angeles, CA, USA, Jun. 2020, pp. 1171–1176.

[3] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection* (Monographs on Statistics and Applied Probability 136). Boca Raton, Fl, USA: CRC Press, Aug. 2014.

[4] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[5] V. V. Veeravalli and T. Banerjee, "Quickest change detection," in *Library in Signal Processing*. Oxford, U.K.: Academic, 2014, pp. 209–255.

[6] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.

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

[8] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Stat.*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.

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

[10] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probab. Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[11] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Probab. Appl.*, vol. 49, no. 3, pp. 458–497, Jan. 2005.

[12] A. G. Tartakovsky and V. V. Veeravalli, "Change-point detection in multichannel and distributed systems," in *Applied Sequential Methodologies: Real-World Examples With Data Analysis*, vol. 173. London, U.K.: Taylor Francis, Jan. 2004, pp. 339–370.

[13] 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 Process.*, vol. 54, no. 9, pp. 3372–3382, Sep. 2006.

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

[15] Y. Mei, "Quickest detection in censoring sensor networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul. 2011, pp. 2148–2152.

[16] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Ann. Stat.*, vol. 41, no. 2, pp. 670–692, Jul. 2013.

[17] G. Fellouris and A. G. Tartakovsky, "Multichannel sequential detection—Part I: Non-I.I.D. data," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4551–4571, Jul. 2017.

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

[19] S. Zou, G. Fellouris, and V. V. Veeravalli, "Quickest change detection under transient dynamics: Theory and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1397–1412, Mar. 2019.

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

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

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

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

[24] R. Zhang and Y. Mei, "Asymptotic statistical properties of communication-efficient quickest detection schemes in sensor networks," *Sequential Anal.*, vol. 37, no. 3, pp. 375–396, 2018.

[25] G. V. Moustakides and V. V. Veeravalli, "Sequentially detecting transitory changes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 26–30.

[26] G. Rovatsos, J. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Statistical power system line outage detection under transient dynamics," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2787–2797, Jun. 2017.

[27] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Quickest change detection under transient dynamics," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 4785–4789.

[28] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Quickest detection of a moving target in a sensor network," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 2399–2403.

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

[30] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, nos. 1–2, pp. 100–115, Jun. 1954.

[31] A. Kott, A. Swami, and B. J. West, "The Internet of Battle Things," *IEEE Comput.*, vol. 49, no. 12, pp. 70–75, Nov. 2016.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Intersci., Jul. 2006.

[33] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Honolulu, Hawaii, May 2007, pp. 317–320.

[34] G. Rovatsos, V. V. Veeravalli, D. Towsley, and A. Swami, "Quickest detection of growing dynamic anomalies in networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, Jul. 2020, pp. 8926–8930.

Georgios Rovatsos (Student Member, IEEE) received the B.S. degree in electrical engineering from the University of Patras, Patras, Greece, in 2014, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana–Champaign in 2016 and 2020, respectively. He currently works as a Senior Data Scientist for Walmart Labs. His research interests are in statistical inference and machine learning.

George V. Moustakides (Life Senior Member, IEEE) received the Diploma degree in electrical and mechanical engineering from the National Technical University of Athens, Greece, in 1979, the M.Sc. degree in systems engineering from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1980, and the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton NJ, USA, in 1983. Since 2007, he has been a Professor with the Department of Electrical and Computer Engineering, University of Patras, Patras, Greece. He also held long-term appointments as a Junior and a Senior Researcher with INRIA, France, as a Professor with the University of Thessaly, Greece, and as a Professor of Practice with Rutgers University, USA. He also held numerous appointments as a Visiting Scholar and an Adjunct Professor with the Princeton University, University of Pennsylvania, Columbia University, University of Maryland, Georgia Institute of Technology, University of Southern California, University of Illinois at Urbana-Champaign, and Rutgers University. His interests include sequential detection, statistical signal processing, and machine learning. From 2011 to 2014, he has served as an Associate Editor for Detection and Estimation and in 2016, as an (inaugural) Associate Editor for Sequential Methods for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Venugopal V. Veeravalli (Fellow, IEEE) received the B.Tech. degree in electrical engineering (Silver Medal Hons.) from the Indian Institute of Technology Bombay, Mumbai, in 1985, the M.S. degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1987, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana–Champaign in 1992. He joined the University of Illinois at Urbana–Champaign in 2000, where he is currently the Henry Magnuski Professor with the Department of Electrical and Computer Engineering, and where he is also affiliated with the Department of Statistics and Coordinated Science Laboratory. Prior to joining the University of Illinois, he was on Faculty of the ECE Department, Cornell University. He served as a Program Director of Communications Research, U.S. National Science Foundation from 2003 to 2005. His research interests span the theoretical areas of statistical inference, machine learning, and information theory, with applications to data science, wireless communications, and sensor networks. Among the awards, he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award, the Presidential Early Career Award for Scientists and Engineers, and the Wald Prize in Sequential Analysis. He has been on the board of governors of the IEEE Information Theory Society. He has been an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY and for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a Distinguished Lecturer for the IEEE Signal Processing Society from 2010 to 2011.