Quickest Detection of Anomalies of Varying Location and Size in Sensor Networks

GEORGIOS ROVATSOS, Student Member, IEEE
VENUGOPAL V. VEERAVALLI, Fellow, IEEE
ECE Department, University of Illinois at Urbana-Champaign Champaign, IL 61801 USA

DON TOWSLEY, Life Fellow, IEEE
Department of Computer Science, University of Massachusetts Amherst, Amherst, MA 01003 USA

ANANTHRAM SWAMI, Life Fellow, IEEE
DEVCOM Army Research Laboratory, Adelphi, MD 20783 USA

The problem of sequentially detecting the emergence of a moving anomaly in a sensor network is studied. In the setting considered, the data-generating distribution at each sensor can alternate between a *nonanomalous* distribution and an *anomalous* distribution. Initially, the observations of each sensor are generated according to its associated nonanomalous distribution. At some *unknown* but *deterministic* time instant, a moving anomaly emerges in the network. It is assumed that the number as well as the identity of the sensors affected by the anomaly may vary with time. While a sensor is affected, it generates observations according to its corresponding anomalous distribution. The goal of this work is to design detection procedures to detect the emergence of such a moving anomaly as quickly as possible, subject to constraints on the frequency of *false alarms*. The problem is studied in a *quickest change detection* framework where it is assumed that the spatial evolution of the anomaly over time is *unknown* but *deterministic*. We modify the worst-path detection delay metric introduced in prior work on moving anomaly detection to consider the case of a moving anomaly of varying size. We then establish that a *weighted dynamic cumulative sum* type test is first-order asymptotically optimal under a delay-false alarm formulation for the proposed worst-path delay as the mean time to false alarm goes to infinity. We conclude by presenting numerical simulations to validate our theoretical analysis.

Manuscript received July 31, 2020; revised April 22, 2021; released for publication April 23, 2021. Date of publication June 11, 2021; date of current version August 9, 2021.

DOI. No. 10.1109/TAES.2021.3088425

This article was presented in part at the 45th IEEE International Conference on Acoustics, Speech, and Signal Processing (online), May 2020 [1]. Refereeing of this contribution was handled by B. Ahmad.

This work was supported in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196. The work of G. Rovatsos and V. V. Veeravalli was supported in part by the National Science Foundation (NSF) under Grant CCF 16-18658

Authors' addresses: Georgios Rovatsos and Venugopal V. Veeravalli are with the ECE Department, University of Illinois at Urbana-Champaign Champaign, IL 61801 USA, E-mail: (rovatso2@illinois.edu; vvv@illinois.edu); Don Towsley is with the Department of Computer Science, University of Massachusetts Amherst, Amherst, MA 01003 USA, E-mail: (towsley@cs.umass.edu); Ananthram Swami is with the DEVCOM Army Research Laboratory, Adelphi, MD 20783 USA, E-mail: (ananthram.swami.civ@mail.mil). *(Corresponding author: Venugopal V. Veeravalli.)*

I. INTRODUCTION

In anomaly detection studied under the *quickest change detection* (QCD) framework, the emergence of an anomaly in the system is assumed to induce a change in the data-generating distribution of the observations obtained by the sensors monitoring the system. The goal is to design a detection algorithm, in the form of a *stopping time*, to detect this change in distribution as quickly as possible, subject to constraints on the frequency of *false alarm* (FA) events. This tradeoff is posed in a stochastic optimization framework, the solution to which depends on the definition of the delay and FA metrics, on the *changepoint* model, as well as on the underlying statistical observation model.

The classical QCD setting involves a sequence of observations undergoing a change from a *nonanomalous* to an *anomalous* distribution. Initially, the observations are *independent and identically distributed* (i.i.d.) according to a known nonanomalous distribution. At some unknown time instant, referred to as the *changepoint*, an anomaly affects the system leading to a persistent change in the data-generating distribution. Thereafter, observations are generated i.i.d. according to a known anomalous distribution. This QCD setting, often referred to as the *i.i.d. model*, has been mostly studied under two formulations: 1) the *minimax* setting [2]–[5], where the changepoint is modeled as *unknown* but *deterministic* and the goal is to minimize a worst average detection delay (WADD) subject to a constraint on the *mean time to false alarm* (MTFA); 2) the *Bayesian* setting [6], [7], where the changepoint is a random variable of known probability distribution and the goal is to minimize an average detection delay subject to constraints on the probability of FA.

QCD theory has also been extensively applied to anomaly detection in sensor networks. In these settings, the resulting QCD problem depends on how the sets of nodes affected by the anomaly change as time progresses. A trivial case arises when the anomaly constantly affects a specific set of sensors, which is known to the decision maker. In this case, the prechange and postchange joint distributions are completely specified and the classical algorithms for QCD [2]–[7] can be extended to provide optimal solutions. A generalized instance of the sensor network anomaly detection problem arises by assuming lack of knowledge of the aforementioned set of nodes affected by the anomaly. This composite postchange model problem has been extensively studied in the asymptotic regime under the minimax setting [8]–[14]. The aforementioned QCD problems can be further generalized by considering an anomaly that affects sensors at different time instants [15]–[23]. As noted up to this point, to our knowledge, sensor network anomaly detection problems are mostly studied in the QCD literature under the core assumption that an anomaly can only affect sensors in a persistent matter.

In this article, we study a problem of sequential moving anomaly detection in sensor networks under the minimax frameworks of Lorden and Pollak [2], [3]. In the setting considered, the sets of nodes affected by the anomaly may vary with time and are unknown to the decision maker. As a result, the anomaly is persistent in the network as a whole without necessarily persistently affecting specific nodes. The problem was initially studied in [24] and [25], where it was assumed that the anomaly evolves according to a *discrete time Markov chain* and is of fixed size. In [26]–[28], the Markov assumption on the anomaly was lifted. Consequently, a worst-path approach was employed where the trajectory of the anomaly was assumed to be *unknown* and *deterministic*. To this end, the lack of a specific model for the anomaly path was counterbalanced by introducing a modification to Lorden's [2] and Pollak's [3] detection delay that evaluates stopping procedures according to their worst performance with respect to the path of the anomaly. Optimal detection procedures were proposed for both Markov chain and worst-path settings for the case of a moving anomaly affecting a set of sensors of constant size.

In this article, we generalize the aforementioned worst-path setting by considering an anomaly of varying size. In particular, we assume that the size of the anomaly varies with time in a series of cascading transient phases, each corresponding to a specific anomaly size before reaching the persistent anomaly size. This problem is particularly relevant to settings where the locations of the sensors are such that different numbers of nodes perceive the anomaly as time progresses or where the physical properties of the anomaly change. For example, consider the case of a target to be detected that approaches a monitored sensor network and, as a result, affects a larger number of nodes with time, eventually entering the region of the network at which point the number of affected sensors remains constant. Another example is the case where multiple targets approach a sensor network at different time instants. The aforementioned settings are often encountered in Internet of Battlefield Things (IoBT) and intrusion detection applications [9], [29].

The analysis in this article draws on the results in [16], where the problem of transient QCD is studied. The main difference between the two works is that, in [16], the joint postchange distribution at each phase is fixed and known to the decision maker. In our work, the joint postchange distribution may vary even during a specific phase since the set of affected nodes changes with time. In addition, the identity of affected nodes is not known to the decision maker; hence, the joint postchange distribution is also not known at any time instant. We establish that a solution to a specific instance of the transient QCD problem studied in [16] also leads to a first-order asymptotically optimal algorithm for the setting studied here.

The rest of this article is organized as follows. In Section II, we introduce necessary notation, describe the underlying statistical model, and present the delay and FA metrics to be used, along with the optimization problem to be solved. In Section III, we introduce our proposed detection algorithm. In Section IV, we present the first-order asymptotic optimality of the proposed test. In Section V, we provide simulation results to numerically validate the use of our proposed detection procedure. Finally, Section VI, concludes this article.

II. PROBLEM MODEL

In this section, we present the observation model corresponding to the moving anomaly detection problem of interest. We also introduce the delay-FA optimization framework to be employed together with the corresponding detection delay metric. We begin by introducing some necessary notation. For any sequence $\{\alpha[k]\}_{k=1}^{\infty}$ and $k_2 > k_1$, we have that $\prod_{j=k_2}^{k_1} \alpha[j] \triangleq 1$ and $\sum_{j=k_2}^{k_1} \alpha[j] \triangleq 0$. Furthermore, $\alpha[k_1, k_2] \triangleq [\alpha[k_1], \ldots \alpha[k_2]]^\top$ denotes the samples from time k_1 to k_2. Given a set E, $|E|$ denotes the number of elements in the set. The set $\{1, \ldots, L\}$ is denoted by $[L]$. Let $\{\boldsymbol{X}[k]\}_{k=1}^{\infty}$ denote the sequence of observations generated by the sensor network, where $\boldsymbol{X}[k] \triangleq [X_1[k], \ldots, X_L[k]]^\top$ denotes the observation vector at time k and $X_\ell[k] \in \mathbb{R}$ denotes the measurement obtained by sensor $\ell \in [L]$ at time k. Define by $\mathscr{F} \triangleq \{\mathscr{F}_k\}_{k=1}^{\infty}$ the filtration generated by the observation process, where $\mathscr{F}_k = \sigma(\boldsymbol{X}[1, k])$ denotes the σ-algebra generated by $\boldsymbol{X}[1, k]$. Finally, for functions $f : \mathbb{R} \mapsto \mathbb{R}$, $g : \mathbb{R} \mapsto \mathbb{R}$, we have that $f(x) \sim g(x)$ denotes that $f(x)/g(x) = 1 + o(1)$ as $x \to \infty$, where $o(1) \to 0$ as $x \to \infty$.

A. Observation Model

Consider a sensor network with sensors given in $[L]$. For sensor $\ell \in [L]$, denote the respective nonanomalous and anomalous distributions by $g_\ell(x)$ and $f_\ell(x)$. Before the emergence of the moving anomaly in the network, it is assumed that sensors generate data i.i.d. across time with respect to their nonanomalous distributions. Furthermore, it is assumed that observations are independent across sensors and time. As a result, the joint probability density function (pdf) of the observations before the anomaly emerges is given by

$$g(\boldsymbol{x}) \triangleq \prod_{\ell=1}^{L} g_\ell(x_\ell).$$

At some *unknown* and *deterministic changepoint* $\nu_1 \geq 1$, a moving anomaly emerges in the network, affecting different sets of sensors as time progresses. It is assumed that the number of affected nodes changes in phases before resolving to a persistent anomaly size. In particular, we assume that our system goes through $K-1$, $K \geq 2$, transient phases before reaching the persistent-size phase, each phase corresponding to a specific moving anomaly size. Phase $i \in [K]$ is assumed to begin at an *unknown* and *deterministic* time instant ν_i, where $\nu_i \geq \nu_{i'}$ for $i > i'$. As a result, the duration of the ith transient phase is given by $d_i \triangleq \nu_{i+1} - \nu_i$ for $i \in [K-1]$. We denote by $\boldsymbol{d} \triangleq \{d_i\}_{i=1}^{K-1}$ the vector containing the transient phase durations. Note that we assume that, in addition to the changepoints, the durations of the transient phases are also *deterministic* and *unknown*. In addition, without loss of generality, we assume that adjacent phases correspond to distinct anomaly sizes. Define by $m^{(i)} \in [L]$ the size of the anomaly at phase $i \in [K]$, which is assumed to be known to the decision maker.

Denote by $\boldsymbol{S}^{(i)} \triangleq \{\boldsymbol{S}^{(i)}[k]\}_{k=1}^{\infty}$ the unknown but deterministic trajectory of the anomaly at phase i, where $\boldsymbol{S}^{(i)}[k]$ denotes the vector containing the anomalous nodes at time k and phase i. Note that $\boldsymbol{S}^{(i)}[k]$ is defined for all $k \geq 1$ and not only $\nu_i \leq k < \nu_{i+1}$ for notational convenience, although only the values at $\nu_i \leq k < \nu_{i+1}$ affect the distribution of our observations. Define by $\mathcal{E}^{(i)}$ the set of vector-values $\boldsymbol{S}^{(i)}[k]$ can take, corresponding to all anomaly positions for an anomaly of size $m^{(i)}$. Note that there are $|\mathcal{E}^{(i)}| = \binom{L}{m^{(i)}}$ such positions (without loss of generality, we assume that the components of each vector are ordered to provide a unique vector per anomaly placement). Assume that the observations are independent across time, *conditioned* on the values of the changepoints ν_i, $i \in [K]$ and on the anomaly trajectory. Then, for a fixed set of trajectory sequences $\boldsymbol{S} \triangleq \{\boldsymbol{S}^{(i)}\}_{i=1}^{K}$ and fixed changepoints $\{\nu_i\}_{i=1}^{K}$, we have that for $i \in [K]$ and $\nu_i \leq k < \nu_{i+1}$ (assuming $\nu_{K+1} \triangleq \infty$)

$$\boldsymbol{X}[k] \sim p_{\boldsymbol{S}^{(i)}[k]}(\boldsymbol{x}) \triangleq \left(\prod_{\ell \in \boldsymbol{S}^{(i)}[k]} f_\ell(x_\ell)\right) \cdot \left(\prod_{\ell \notin \boldsymbol{S}^{(i)}[k]} g_\ell(x_\ell)\right)$$

where for vector $\boldsymbol{E}$, $p_{\boldsymbol{E}}(\boldsymbol{x})$ denotes the joint pdf induced on a vector observation when the identities of the anomalous nodes are contained in $\boldsymbol{E}$. As a result, *conditioned* on $\{\nu_i\}_{i=1}^{K}$ and $\boldsymbol{S}$, the complete statistical model is

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{x}), & 1 \leq k < \nu_1 \\ p_{\boldsymbol{S}^{(i)}[k]}(\boldsymbol{x}), & \nu_i \leq k < \nu_{i+1} \end{cases} \tag{1}$$

$i \in [K]$, where we assume that the observations are independent across time, conditioned on the changepoints.

B. Delay-FA Tradeoff Formulation

Our goal in this work is to design a detection algorithm to detect the abrupt distribution change occurring at time ν_1 as described in (1) as quickly as possible, subject to FA constraints. To this end, we frame this detection problem in a QCD setting [30]–[32], with detection procedures taking the form of *stopping times*. A stopping time τ adapted to $\mathscr{F}$ is a positive random variable which satisfies $\{\tau \leq k\} \in \mathscr{F}_k$ for all $k \geq 1$, i.e., the decision to raise an alarm at time k is determined only by the observations up to that point. To frame the aforementioned delay-FA tradeoff, we introduce a modified version of Lorden's delay metric [2] to account for the lack of anomaly path knowledge. In particular, we evaluate candidate stopping times according to the trajectory of the anomaly that leads to the anomaly inducing the largest detection delay. Explicitly, denote by $\mathbb{E}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}[\cdot]$ the expectation under the statistical model in (1) for fixed ν_1, $\boldsymbol{d}$, and $\boldsymbol{S}$. Then, for any stopping rule τ adapted to $\mathscr{F}$ and for vector $\boldsymbol{d}$, define the following modification of Lorden's [2] WADD metric:

$$\begin{aligned}&\mathrm{WADD}_{\boldsymbol{d}}(\tau)\\ &= \sup_{\boldsymbol{S}} \sup_{\nu_1 \geq 1} \operatorname{ess\,sup} \mathbb{E}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}\left[\tau - \nu_1 + 1 | \mathscr{F}_{\nu_1 - 1}, \tau \geq \nu_1\right]\end{aligned}$$

where $\mathbb{E}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}[\tau - \nu_1 + 1 | \mathscr{F}_{\nu_1-1}, \tau \geq \nu_1] \triangleq 1$ when $\mathbb{P}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}(\tau \geq \nu_1) = 0$ by convention. Note that the main difference from Lorden's [2] classical detection delay metric is the use of an additional sup over the path of the anomaly. Furthermore, note that the proposed detection delay depends on $\boldsymbol{d}$ since different phase durations imply different probability distributions across time and, hence, different algorithm performance. Denote by $\mathbb{E}_\infty[\cdot]$ the expectation when no anomaly is present in the network. To quantify the frequency of FA events, we use the MTFA, denoted by $\mathbb{E}_\infty[\tau]$ for stopping time τ. For $\gamma > 1$, a predetermined constant, define the class of stopping times

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

Our goal is to design a stopping procedure τ that solves the following stochastic optimization problem:

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

for any value of $\boldsymbol{d}$.

C. Randomized Anomaly Allocation Model

In this subsection, we introduce an alternative statistical model to that in (1). Note that this model is only used as an intermediate tool that will play an important role in the presentation of our results as well as in the analysis and that it is not the model characterizing our QCD problem. More explicitly, consider the case of a moving anomaly that, at each phase i, affects one of the sets of sensors in $\mathcal{E}^{(i)}$ at random. To this end, denote by $\boldsymbol{\alpha}^{(i)} \triangleq \{\alpha_{\boldsymbol{E}}^{(i)} : \boldsymbol{E} \in \mathcal{E}^{(i)}\}$ the *probability mass function* (pmf) containing the probabilities that each of the vectors in $\mathcal{E}^{(i)}$ is chosen as the vector of anomalous nodes at each time instant at phase i. In particular, at each time instant in phase i, the anomalous nodes are chosen i.i.d. from $\mathcal{E}^{(i)}$ according to $\boldsymbol{\alpha}^{(i)}$. Define $\boldsymbol{\alpha} \triangleq \{\boldsymbol{\alpha}^{(i)}\}_{i=1}^{K}$ to be the set of the aforementioned pmfs for all phases. According to this randomized allocation model, the joint pdf before the emergence of the anomaly is identical to that in (1). In addition, after the emergence of the anomaly, we have that the joint pdf of the observations at phase i is completely specified by

$$\overline{p}_{\boldsymbol{\alpha}^{(i)}}^{(i)}(\boldsymbol{x}) \triangleq \sum_{\boldsymbol{E} \in \mathcal{E}^{(i)}} \alpha_{\boldsymbol{E}}^{(i)} p_{\boldsymbol{E}}(\boldsymbol{x}). \tag{2}$$

For fixed $\{\nu_i\}_{i=1}^{K-1}$, $\boldsymbol{\alpha}$, this results in the following statistical observation model:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{x}), & 1 \leq k < \nu_1 \\ \overline{p}_{\boldsymbol{\alpha}^{(i)}}^{(i)}(\boldsymbol{x}), & \nu_i \leq k < \nu_{i+1} \end{cases} \tag{3}$$

$i \in [K]$. Furthermore, for fixed $\boldsymbol{\alpha}$, define the Kullback–Leibler (KL) divergence [33] between the joint pdf at phase i and the nonanomalous joint pdf $g(\boldsymbol{x})$ by

$$I^{(i)}_{\boldsymbol{\alpha}^{(i)}} \triangleq D(\overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}} \| g) \triangleq \mathbb{E}_{\overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}}}\left[\log \frac{\overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}}(\boldsymbol{X})}{g(\boldsymbol{X})}\right] \tag{4}$$

where, for pdf $p(\boldsymbol{x})$, $\mathbb{E}_p[\cdot]$ denotes the expectation with respect to $p(x)$. Note that (3) corresponds to the observation model for a transient QCD problem, as described in [16] since the prechange and postchange pdfs are completely specified. This QCD problem is associated with a corresponding detection delay. In particular, let $\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu_1,\boldsymbol{d}}[\cdot]$ denote the expectation under the model in (3) for fixed ν_1, $\boldsymbol{d}$, and $\boldsymbol{\alpha}$. Then, for stopping time τ, define the detection delay corresponding to the QCD problem detailed in (3) by

$$\overline{\mathrm{WADD}}_{\boldsymbol{\alpha},\boldsymbol{d}}(\tau) = \sup_{\nu_1 \geq 1} \operatorname{ess\,sup} \overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu_1,\boldsymbol{d}}[\tau - \nu_1 + 1 | \mathscr{F}_{\nu_1 - 1}, \tau \geq \nu_1] \tag{5}$$

where the convention that $\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,\boldsymbol{d}}[\tau - \nu_1 + 1 | \mathscr{F}_{\nu_1 - 1}, \tau \geq \nu_1] \triangleq 1$ when $\overline{\mathbb{P}}^{\boldsymbol{\alpha}}_{\nu_1,\boldsymbol{d}}(\tau \geq \nu_1) = 0$ is also used here. Note that the transient QCD problem described in (3)–(5) can be solved asymptotically by using the weighted dynamic cumulative sum (WD-CUSUM) test proposed in [16] adapted to the model in (3). However, it is not clear whether this solution coincides with that of Section II-B. In the remainder of the article, we show that solving the transient QCD problem in (3)–(5) for a specific choice of pmfs in $\boldsymbol{\alpha}$ will lead to the solution of the initial worst-path problem described in Section II-B.

III. PROPOSED DETECTION ALGORITHM

In this section, we present the mixture-WD-CUSUM (M-WD-CUSUM) test that solves the transient QCD problem introduced in (3)–(5). In particular, for a fixed set of pmfs $\boldsymbol{\alpha}$, define the log-likelihood ratio at time k during phase i by

$$Z^{(i)}_{\boldsymbol{\alpha}}[k] = \log \frac{\overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}}(\boldsymbol{X}[k])}{g(\boldsymbol{X}[k])}. \tag{6}$$

Consider the following M-WD-CUSUM test statistic

$$W_{\boldsymbol{\alpha}}[k] = \max\{\Omega^{(1)}_{\boldsymbol{\alpha}}[k], \ldots, \Omega^{(K)}_{\boldsymbol{\alpha}}[k], 0\} \tag{7}$$

where for $i \in [K]$, $\Omega^{(i)}_{\boldsymbol{\alpha}}[k]$ is calculated recursively as

$$\Omega^{(i)}_{\boldsymbol{\alpha}}[k] = \max_{0 \leq j \leq i} \left(\Omega^{(j)}_{\boldsymbol{\alpha}}[k-1] + \sum_{r=j}^{i-1} \log \rho_r \right) + Z^{(i)}_{\boldsymbol{\alpha}}[k] + \log(1 - \rho_i) \tag{8}$$

where $\rho_0 \triangleq 1$, $\rho_i \in (0, 1)$ for $i \in [K-1]$, $\rho_K \triangleq 0$, $\Omega^{(i)}[0] \triangleq 0$ for all $i \in [K]$, and $\Omega^{(0)}[k] \triangleq 0$ for all k. Furthermore, define the corresponding stopping time by

$$\tau_W(\boldsymbol{\alpha}, b) \triangleq \inf\{k \geq 1 : W_{\boldsymbol{\alpha}}[k] \geq b\}. \tag{9}$$

From the results in [16], the M-WD-CUSUM test presented in (7)–(9) is asymptotically optimal with respect to the transient QCD problem in (3)–(5) for specific choices of ρ_i parameters. Explicitly, the ρ_i parameters are introduced so that the FA constraint is satisfied for $b = \log \gamma$ and should be chosen to not play a role asymptotically in order for an optimal test to be derived. More details regarding the choice of the ρ_i parameters will be given in the subsequent analysis and can also be found in [16].

IV. ASYMPTOTIC OPTIMALITY OF THE M-WD-CUSUM PROCEDURE

In this section, we establish the asymptotic optimality of the M-WD-CUSUM test for a carefully chosen $\boldsymbol{\alpha}$.

A. Universal Asymptotic Lower Bound on the WADD

We begin our analysis by presenting an asymptotic lower bound on WADD for stopping times in $\mathcal{C}_\gamma$. Our lower bound is based on an important lemma connecting WADD and $\overline{\mathrm{WADD}}$. In particular, note that WADD evaluates each candidate stopping time τ with regard to the worst possible path of the anomaly for τ. On the other hand, $\overline{\mathrm{WADD}}$ corresponds to a model where the anomalous nodes are chosen at random. Our first result says that the worst-path delay cannot be smaller than the delay that corresponds to choosing the anomalous nodes at random regardless of the choice of prior $\boldsymbol{\alpha}$. In particular, we have the following lemma.

LEMMA 1 For any stopping time τ, any vector of pmfs $\boldsymbol{\alpha}$, and any $\boldsymbol{d}$, we have that

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau) \geq \overline{\mathrm{WADD}}_{\boldsymbol{\alpha},\boldsymbol{d}}(\tau).$$

PROOF See Appendix. ■

Since the results in [16] provide a universal asymptotic lower bound on $\overline{\mathrm{WADD}}$ for any $\boldsymbol{\alpha}$, an asymptotic lower bound on WADD then follows directly from Lemma 1. However, since the asymptotic rate in the lower bound of $\overline{\mathrm{WADD}}$ is a function of the KL numbers defined in (4), we need to choose the pmfs in $\boldsymbol{\alpha}$ to get the tightest lower bound on WADD. To this end, define

$$\boldsymbol{\alpha}^{(i)}_* \triangleq \arg\min_{\boldsymbol{\alpha}^{(i)}} I^{(i)}_{\boldsymbol{\alpha}^{(i)}} \tag{10}$$

and $\boldsymbol{\alpha}_* \triangleq \{\boldsymbol{\alpha}^{(i)}_*\}_{i=1}^K$, the vector containing the minimizing pmfs. It can be shown that $I^{(i)}_{\boldsymbol{\alpha}^{(i)}}$ is strictly convex with respect to $\boldsymbol{\alpha}$; hence, such a minimizer is uniquely defined. Furthermore, define the minimum value of $I^{(i)}_{\boldsymbol{\alpha}^{(i)}}$ by

$$I^{(i)}_* \triangleq I^{(i)}_{\boldsymbol{\alpha}^{(i)}_*}.$$

To ensure that the transient phases play a nontrivial role asymptotically in our initial QCD problem, the durations of the transient phases need to scale to infinity with γ. In particular, without loss of generality, assume that there exist constants $c_i \in [0, \infty) \cup \{\infty\}$, $i \in [K-1]$ such that as $\gamma \to \infty$

$$d_i \sim c_i \frac{\log \gamma}{I^{(i)}_*}. \tag{11}$$

This assumption can be intuitively explained since asymptotically the rate of the transient durations with respect to $\log\gamma$ will indicate the phase at which the anomaly will be detected [16]. The specific choice of KL numbers in the scaling coefficient is chosen so that the universal lower bound and upper bound on the delay of the proposed test match, as will be noted in the upper bound analysis. We then have the following:

THEOREM 1 Assume that (11) holds. Furthermore, define $h \triangleq \min\{j \in [K] : \sum_{i=1}^{j} c_i \geq 1\}$. We then have that as $\gamma \to \infty$

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}_{\boldsymbol{d}}(\tau) \geq \log\gamma \left(\sum_{i=1}^{h-1} \frac{c_i}{I_*^{(i)}} + \frac{1 - \sum_{i=1}^{h-1} c_i}{I_*^{(h)}} \right) \cdot (1 - o(1)).$$

PROOF The result follows directly by applying Lemma 1 for $\boldsymbol{\alpha} = \boldsymbol{\alpha}_*$ and [16, Theorem 5] to lower bound $\overline{\mathrm{WADD}}_{\boldsymbol{\alpha}_*,\boldsymbol{d}}(\tau)$. ■

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

We now establish an asymptotic upper bound on the WADD of the proposed algorithm. The asymptotic upper bound is based on exploiting the upper bound analysis in [16], [27], and [28]. For the asymptotic upper bound analysis to be nontrivial, we need to assume that the transient durations scale accordingly to threshold b. In particular, assume that there exist constants $c_i' \in [1, \infty) \cup \{\infty\}$, $i \in [K-1]$ such that

$$d_i \sim c_i' \frac{b}{I_*^{(i)}}. \tag{12}$$

Furthermore, to use the analysis in [16], we need to design the parameters ρ_i, $i \in [K-1]$ in the M-WD-CUSUM such that their effect is asymptotically negligible [16]. In particular, assume that ρ_i can be chosen such that

$$\rho_i \to 0, \text{ and } -\frac{\log\rho_i}{b} \to 0 \tag{13}$$

as $b \to \infty$ for $i \in [K-1]$. We then have the following asymptotic upper bound.

THEOREM 2 Suppose b and ρ_i, $i \in [K-1]$ are chosen such that (12) and (13) hold. Assume that

$$\max_{i \in [K]} \max_{\boldsymbol{E} \in \mathcal{E}^{(i)}} \mathbb{E}_{p_{\boldsymbol{E}}} \left[\left(\log \frac{\overline{p}^{(i)}_{\boldsymbol{\alpha}_*^{(i)}}(\boldsymbol{X})}{g(\boldsymbol{X})} \right)^2 \right] < \infty. \tag{14}$$

Furthermore, define $h' \triangleq \min\{j \in [K] : \sum_{i=1}^{j} c_i' \geq 1\}$. We then have that as $b \to \infty$

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau_W(\boldsymbol{\alpha}_*, b)) \leq b \left(\sum_{i=1}^{h-1} \frac{c_i'}{I_*^{(i)}} + \frac{1 - \sum_{i=1}^{h'-1} c_i'}{I_*^{(h')}} \right) \cdot (1 + o(1)).$$

PROOF See Appendix. ■

C. Asymptotic Optimality of the M-WD-CUSUM Test

By combining Theorem 1 with Theorem 2, we can establish the asymptotic optimality of the M-WD-CUSUM when $\boldsymbol{\alpha} = \boldsymbol{\alpha}_*$. In particular, we have the following theorem.

THEOREM 3 Assume that

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

We then have the following:

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

2) Assume that $\boldsymbol{d}$ is chosen to satisfy (11) as $\gamma \to \infty$ for some $c_i \in [1, \infty) \cup \{\infty\}$, $i \in [K-1]$ and that as $\gamma \to \infty$

$$\rho_i \to 0, \text{ and } -\frac{\log\rho_i}{\log\gamma} \to 0.$$

Let $h \triangleq \min\{j \in [K] : \sum_{i=1}^{j} c_i \geq 1\}$. We then have that as $\gamma \to \infty$

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau_W(\boldsymbol{\alpha}_*, \log\gamma)) \sim \inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}_{\boldsymbol{d}}(\tau) \sim \log\gamma \left(\sum_{i=1}^{h-1} \frac{c_i}{I_*^{(i)}} + \frac{1 - \sum_{i=1}^{h-1} c_i}{I_*^{(h)}} \right).$$

PROOF 1) Follows directly from the MTFA analysis of the WD-CUSUM test [16]. ■

2) Follows from 1) and Theorems 1 and 2, and because $b = \log\gamma$, we have that $c_i = c_i'$ for all $i \in [K-1]$.

REMARK Note that the first-order asymptotic optimality result in this article also holds if we use a worst-path version of Pollak's detection delay [3]. In particular, for stopping time τ and vector of transient durations $\boldsymbol{d}$, define the detection delay

$$\mathrm{CADD}_{\boldsymbol{d}}(\tau) \triangleq \sup_{\boldsymbol{S}} \sup_{\nu_1 \geq 1} \mathbb{E}^{\boldsymbol{S}}_{\nu_1,\boldsymbol{d}} \left[\tau - \nu_1 | \tau \geq \nu_1 \right].$$

By deriving a lower bound similar to the one in Lemma 1 and since WADD is always larger than CADD, we can easily establish the first-order asymptotic optimality of the M-WD-CUSUM test under Pollak's criterion, i.e., Theorem 3 also holds when WADD is replaced by CADD.

V. NUMERICAL RESULTS

In this section, we numerically evaluate the performance of the proposed M-WD-CUSUM algorithm of (6)–(9). We consider the case of both homogeneous and heterogeneous sensors. For the case of homogeneous sensors, it can be shown that the optimal weight $\boldsymbol{\alpha}$ choice in (6)–(9) is uniform [1]. Note that $\mathrm{WADD}_{\boldsymbol{d}}$ for the proposed test is attained at $\nu_1 = 1$, i.e., $\nu_1 = 1$ leads to the worst-case delay. Furthermore, for the case of heterogeneous sensors, the worst-path cannot be specified analytically. As a result, we will approximate the worst-path delay by placing the anomalous nodes at each phase such that the worst-possible slope for the test statistic is attained. This is done because, asymptotically, the performance of the proposed algorithm is dominated by the slope of the test statistic. We numerically calculate the

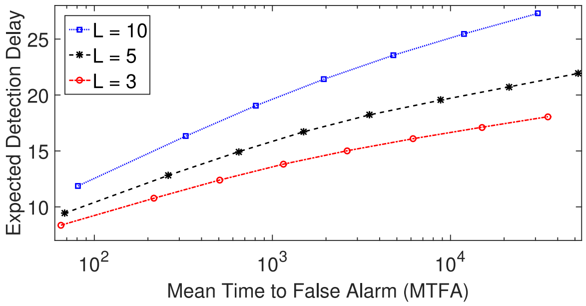


Fig. 1. WADD$_d$ vs. MTFA for $K = 3$ and varying network sizes.

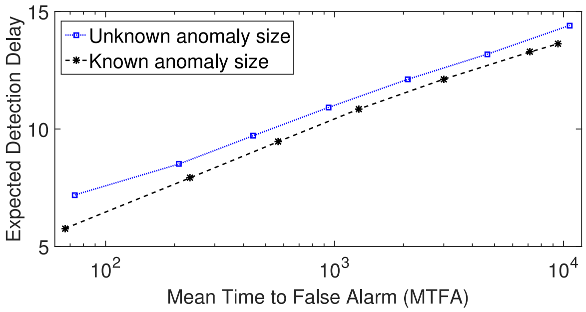


Fig. 2. WADD$_d$ vs. MTFA comparison between the test that exploits and the test that does not exploit complete knowledge of the anomaly size across phases for a homogeneous sensor network.

average statistic slope through Monte Carlo simulations. In addition, we use $\rho_i = \frac{1}{b}$ to guarantee that the conditions in (13) are satisfied.

For the case of homogeneous sensors, we focus on the case of $g_\ell = \mathcal{N}(0, 1)$ and $f_\ell = \mathcal{N}(1, 1)$, $\ell \in [L]$. In Fig. 1, we simulate the proposed M-WD-CUSUM test for the case of $K = 3$, $m^{(1)} = 1$, $m^{(2)} = 2$, $m^{(3)} = 3$, $d_1 = 9$, $d_2 = 10$, and for $L = 3, 5, 10$. We note that for fixed MTFA, the average detection delay increases with network size. This is to be expected since a larger network introduces more noise in the calculation of the mixture likelihood ratios in (8). Furthermore, we see that as the MTFA increases, the slopes of the curves gradually decrease. This means that the M-WD-CUSUM is adaptive to each transient phase since the expected slope increases as the anomaly size increases.

In Fig. 2, we evaluate the performance loss that our algorithm incurs when the anomaly size is not completely specified. In particular, we consider the case of $K = 3$, $m^{(1)} = 2$, $m^{(2)} = 3$, $m^{(3)} = 4$, $d_1 = 9$, $d_2 = 10$, and $L = 6$ and compare the performance of the M-WD-CUSUM test that is designed by completely knowing the values of these parameters with the M-WD-CUSUM that assumes that $K = 6$ and $m^{(i)} = i$ for $i \in [K]$. As expected, the algorithm that exploits complete knowledge of the size of the anomaly at each phase performs much better. Note that the performance loss for our case study is not significant; however, the performance loss can increase significantly as L increases if our estimates for K and $m^{(i)}$ are not sufficiently accurate.

Finally, in Fig. 3, we evaluate the performance of our proposed detection procedure for the case of a heterogeneous sensor network with $L = K = 5$, $m^{(i)} = i$ for $i \in [K]$, $g_\ell = \mathcal{N}(0, 1)$, and $f_\ell = \mathcal{N}(\mu_\ell, 1)$ where $\boldsymbol{\mu} = [0.8, 0.8, 1, 1.2, 1.2]^\top$. Furthermore, we assume that $d_1 = 19$ and $d_2 = d_3 = d_4 = 20$. To this end, we compare the M-WD-CUSUM that uses complete knowledge of $f_\ell(x)$ and $g_\ell(x)$ for all $\ell \in [L]$ and chooses the proposed optimal weights (see Theorems 1–3) to the M-WD-CUSUM test that uses uniform weights (i.e., assumes sensors are homogeneous). For each phase, the anomaly for the uniform weights case is placed so that the slope of the statistic is minimized. We see that there is significant performance loss when the decision maker assumes that the sensors are homogeneous when they are heterogeneous.

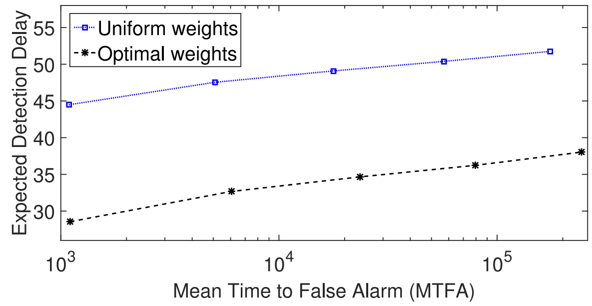


Fig. 3. WADD$_d$ vs. MTFA comparison between the test that exploits and the test that does not exploit complete knowledge of the sensor pdfs for a heterogeneous sensor network.

VI. CONCLUSION

In this article, we studied the problem of sequentially detecting a moving anomaly of varying size in a sensor network. We posed the problem within Lorden's minimax framework, under the assumption that the path of the anomaly is *unknown* but *deterministic*. To account for the lack of a model of how the anomaly evolves spatialy over time, we introduced a modified version of Lorden's [2] delay metric that evaluates candidate stopping rules with respect to the worst-path performance. We proposed a WD-CUSUM-type test that is first-order asymptotically optimal for the case of observations that are independent across time and derived its asymptotic delay rate. It should be noted that our detection procedure can be modified by using conditional pdfs in place of the marginal pdfs in the likelihood ratio of the test statistic in (6)–(9) to yield a first-order asymptotically optimal test even when the observations are dependent across time. However, the analysis in this setting becomes significantly more challenging (see, e.g., [5] and [7]).

Future work in this area includes extending the proposed detection procedure to allow for partial knowledge of the postchange model as well as studying the problem when an adversary can mask the emergence of the anomaly. The first problem is particularly important for the proposed procedures to be applicable in practice. We believe that the algorithms given in this article can be extended to address partial postchange model knowledge by introducing online changepoint as well as data-generating distribution estimators (see, e.g., [34] and [35]).

APPENDIX

For our theoretical analysis, we focus on the case of two postchange phases (one transient phase and one persistent phase). The results in this article hold for the case of arbitrary number of phases K known to the decision maker, but, in that case, the analysis becomes cumbersome. To this end, consider the sequences $\boldsymbol{S}^{(1)} \triangleq \{\boldsymbol{S}^{(1)}[k]\}_{k=1}^{\infty}$ and $\boldsymbol{S}^{(2)} \triangleq \{\boldsymbol{S}^{(2)}[k]\}_{k=1}^{\infty}$ that characterize the location of the anomalous nodes at each time instant for postchange phases 1 and 2, respectively. For clarity of notation, we will use ν in the Appendix to denote the first changepoint ν_1, d to denote the transient duration d_1, and ρ to denote ρ_1. Define the likelihood ratio of samples $\boldsymbol{X}[1,k]$ between the hypothesis that the anomaly evolves according to $\boldsymbol{S}$ and changepoints are equal to ν_1 and ν_2 and the hypothesis that the anomaly never appears by

$$\Gamma_{\boldsymbol{S}}(k,\nu_1,\nu_2) \triangleq \begin{cases} \left[\prod_{j=\nu_1}^{\min\{\nu_2-1,k\}} \prod_{\ell\in \boldsymbol{S}_1[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])}\right] \\ \quad \cdot \left[\prod_{j=\nu_2}^{k} \prod_{\ell\in \boldsymbol{S}_2[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])}\right], & \nu_1<\nu_2, \\ \prod_{j=\nu_1}^{k} \prod_{\ell\in \boldsymbol{S}_2[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])}, & \nu_1=\nu_2. \end{cases}$$

In addition, for the model in (3)–(5), define the likelihood ratio of samples $\boldsymbol{X}[1,k]$ between the hypothesis that the anomaly evolves according to mixture weights in $\boldsymbol{\alpha}$ and changepoints are equal to ν_1 and ν_2 and the hypothesis that the anomaly never appears by

$$\mathcal{L}_{\boldsymbol{\alpha}}(k,\nu_1,\nu_2) \triangleq \begin{cases} \left[\prod_{j=\nu_1}^{\min\{\nu_2-1,k\}} \frac{\overline{p}^{(1)}_{\boldsymbol{\alpha}^{(1)}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}\right] \\ \quad \cdot \left[\prod_{j=\nu_2}^{k} \frac{\overline{p}^{(2)}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}\right], & \nu_1<\nu_2, \\ \prod_{j=\nu_1}^{k} \frac{\overline{p}^{(2)}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}, & \nu_1=\nu_2. \end{cases}$$

We now proceed to establish Lemma 1 and Theorem 2 which imply the main theoretical results of this article. The analysis is based on techniques in [5], [16], and [28].

PROOF OF LEMMA 1 For any stopping rule τ, define its truncated version by $\tau^{(N)} \triangleq \min\{\tau, N\}$ where N is a positive integer. Since $\tau \geq \tau^{(N)}$, it can be established that for any ν, $d, N \geq 1$

$$\begin{aligned} \mathrm{WADD}_d(\tau) &\geq \mathrm{WADD}_d(\tau^{(N)}) \\ &\geq \sup_{\boldsymbol{S}} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)}-\nu+1|\mathscr{F}_{\nu-1},\tau^{(N)}\geq \nu]. \end{aligned} \tag{15}$$

In addition, we have that

$$\begin{aligned} &\mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)}-\nu+1|\mathscr{F}_{\nu-1},\tau^{(N)}\geq \nu] \\ &= \mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &\overset{(a)}{=} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\sum_{j=\nu}^{N} \mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &= \sum_{j=\nu}^{N} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &\overset{(b)}{=} \sum_{j=\nu}^{N} \mathbb{E}_{\infty}\left[\Gamma_{\boldsymbol{S}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &= \mathbb{E}_{\infty}\left[\sum_{j=\nu}^{N}\Gamma_{\boldsymbol{S}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &= \mathbb{E}_{\infty}\left[\Gamma_{\boldsymbol{S}}\left(\nu-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq \nu\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &+\mathbb{E}_{\infty}\left[\sum_{j=\nu+1}^{N}\Gamma_{\boldsymbol{S}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &= \mathbb{E}_{\infty}\left[\mathbb{1}_{\{\tau^{(N)}\geq \nu\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &+\mathbb{E}_{\infty}\left[\sum_{j=\nu+1}^{N}\Gamma_{\boldsymbol{S}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &\overset{(c)}{=} \mathbb{E}_{\infty}\left[\mathbb{1}_{\{\tau^{(N)}\geq \nu\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu]\right] \\ &+\mathbb{E}_{\infty}\left[\sum_{j=\nu}^{N-1}\Gamma_{\boldsymbol{S}}\left(j,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}> j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \end{aligned} \tag{16}$$

where (a) follows since $\mathbb{1}_{\{\tau^{(N)}\geq j\}}=0$ for $j>N$ because $\tau^{(N)}\leq N$, (b) follows from a change of measure, and (c) from a change of variables. As a result, by taking the supremum over $\boldsymbol{S}$, we have that

$$\begin{aligned} &\sup_{\boldsymbol{S}} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)}-\nu+1|\mathscr{F}_{\nu-1},\tau^{(N)}\geq \nu] \\ &= \mathbb{E}_{\infty}\left[\mathbb{1}_{\{\tau^{(N)}\geq \nu\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &+\sup_{\boldsymbol{S}}\mathbb{E}_{\infty}\left[\sum_{j=\nu}^{N-1}\Gamma_{\boldsymbol{S}}\left(j,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}> j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right]. \end{aligned} \tag{17}$$

By bounding the supremum by an average, as in [1] and [28], it can be shown that

$$\begin{aligned} &\sup_{\boldsymbol{S}}\mathbb{E}_{\infty}\left[\sum_{j=\nu}^{N-1}\Gamma_{\boldsymbol{S}}\left(j,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}> j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right] \\ &\geq \mathbb{E}_{\infty}\left[\sum_{j=\nu}^{N-1}\mathcal{L}_{\boldsymbol{\alpha}}\left(j,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}> j\}} \middle| \mathscr{F}_{\nu-1},\tau^{(N)}\geq\nu\right]. \end{aligned} \tag{18}$$

From (15), (17), and (18), and by following similar steps as in (16), we then have that

$$\begin{aligned}
\mathrm{WADD}_d(\tau) &\geq \mathbb{E}_\infty\left[\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&+\mathbb{E}_\infty\left[\sum_{j=\nu}^{N-1}\mathcal{L}_{\boldsymbol{\alpha}}\left(j,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}>j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\mathbb{E}_\infty\left[\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&+\mathbb{E}_\infty\left[\sum_{j=\nu+1}^{N}\mathcal{L}_{\boldsymbol{\alpha}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\mathbb{E}_\infty\left[\sum_{j=\nu}^{N}\mathcal{L}_{\boldsymbol{\alpha}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\sum_{j=\nu}^{N}\mathbb{E}_\infty\left[\mathcal{L}_{\boldsymbol{\alpha}}\left(j-1,\nu,\nu+d\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\sum_{j=\nu}^{N}\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\mathbb{1}_{\{\tau^{(N)}\geq j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\sum_{j=\nu}^{N}\mathbb{1}_{\{\tau^{(N)}\geq j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\sum_{j=\nu}^{\infty}\mathbb{1}_{\{\tau^{(N)}\geq j\}}\Big|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu\right]\\
&=\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}[\tau^{(N)}-\nu+1|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu].
\end{aligned}$$

From the monotone convergence theorem, since $\tau^{(N)}-\nu+1$ and $\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}$ are nondecreasing with N, we have that

$$\begin{aligned}
&\lim_{N\to\infty}\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}[\tau^{(N)}-\nu+1|\mathscr{F}_{\nu-1}, \tau^{(N)}\geq\nu]\\
&=\lim_{N\to\infty}\frac{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[(\tau^{(N)}-\nu+1)\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}|\mathscr{F}_{\nu-1}\right]}{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}|\mathscr{F}_{\nu-1}\right]}\\
&=\frac{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\lim_{N\to\infty}(\tau^{(N)}-\nu+1)\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}|\mathscr{F}_{\nu-1}\right]}{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\lim_{N\to\infty}\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}|\mathscr{F}_{\nu-1}\right]}\\
&=\frac{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[(\tau-\nu+1)\mathbb{1}_{\{\tau\geq\nu\}}|\mathscr{F}_{\nu-1}\right]}{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\mathbb{1}_{\{\tau\geq\nu\}}|\mathscr{F}_{\nu-1}\right]}\\
&=\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\tau-\nu+1|\mathscr{F}_{\nu-1},\tau\geq\nu\right].
\end{aligned}$$

As a result, by taking the sup over ν and the esssup, we have that for any stopping time τ, $\boldsymbol{\alpha}$, and for $d\geq 0$

$$\mathrm{WADD}_d(\tau)\geq\overline{\mathrm{WADD}}_{\boldsymbol{\alpha},d}(\tau).$$

Proof of Theorem 2 Our upper bound analysis is based on the proof technique in [16]. In particular, due to the Markov property and recursive structure of the M-WD-CUSUM test, we have that for any $\boldsymbol{S}$ and any values of b, $\boldsymbol{\alpha}$, and d

$$\mathrm{WADD}_d(\tau_W(\boldsymbol{\alpha},b))=\sup_{\boldsymbol{S}}\mathbb{E}^{\boldsymbol{S}}_{1,d}\left[\tau_W(\boldsymbol{\alpha},b)\right].$$

Furthermore, since $\rho\to 0$ and $-\frac{\log\rho}{b}\to 0$ as $b\to\infty$ and since $d\sim c_1'\frac{b}{I_*^{(1)}}$ as $b\to\infty$, we have that as $b\to\infty$

$$d\sim c_1'\frac{b}{I_*^{(1)}+\log(1-\rho)}.$$

Depending on the value of c_1', we can proceed to bound $\sup_{\boldsymbol{S}}\mathbb{E}^{\boldsymbol{S}}_{1,d}[\tau_W(\boldsymbol{\alpha}_*,b)]$ as in [16].

Case 1: Consider the case of $c_1'>1$. Let $\delta>0$. Choose $\epsilon>0$ such that $1\leq\frac{1+\epsilon}{1-\epsilon}\leq c_1'$ which in turn implies that $\frac{c_1'(1-\epsilon)}{1+\epsilon}\geq 1$ and define

$$\begin{aligned}
n_b &\triangleq\frac{b}{I_*^{(1)}+\log(1-\rho)-\epsilon},\\
c_\epsilon &\triangleq\lfloor c_1'\frac{1-\epsilon}{1+\epsilon}\rfloor.
\end{aligned}$$

We then have that

$$\begin{aligned}
&\sup_{\boldsymbol{S}}\mathbb{E}^{\boldsymbol{S}}_{1,d}\left[\frac{\tau_W(\boldsymbol{\alpha}_*,b)}{n_b}\right]\overset{(a)}{=}\sup_{\boldsymbol{S}}\int_0^\infty\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\tau_W(\boldsymbol{\alpha}_*,b)}{n_b}>x\right)dx\\
&\overset{(b)}{\leq}\sup_{\boldsymbol{S}}\sum_{\zeta=0}^{\infty}\mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_W(\boldsymbol{\alpha}_*,b)>\zeta n_b)\\
&\leq 1+\sum_{\zeta=1}^{c_\epsilon}\sup_{\boldsymbol{S}}\mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_W(\boldsymbol{\alpha}_*,b)>\zeta n_b)\\
&+\lim_{\xi\to\infty}\sum_{\zeta=c_\epsilon+1}^{\xi}\sup_{\boldsymbol{S}}\mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_W(\boldsymbol{\alpha}_*,b)>\zeta n_b)
\end{aligned}\tag{19}$$

where (a) follows by writing the expectation as an integral of the inverse cumulative density function for a positive random variable and (b) from the sum-integral inequality.

We now consider two cases depending on the value of ζ relative to c_ϵ. First, fix $\zeta\in[c_\epsilon]$. We then have that datapoints $\boldsymbol{X}[1,\zeta n_b]$ are all generated in postchange phase 1. As a result, we have that for any $\boldsymbol{S}$ and $\zeta\in[c_\epsilon]$

$$\begin{aligned}
&\mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_W(\boldsymbol{\alpha}_*,b)>\zeta n_b)=\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\max_{1\leq k\leq\zeta n_b}W_{\boldsymbol{\alpha}_*}[k]<b\right)\\
&\overset{(c)}{\leq}\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\sum_{j=(r-1)n_b+1}^{rn_b}\left(Z^{(1)}_{\boldsymbol{\alpha}_*}[j]+\log(1-\rho)\right)<b,\ \forall\, r\in[\zeta]\right)\\
&\overset{(d)}{=}\prod_{r=1}^{\zeta}\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{1}{n_b}\sum_{j=(r-1)n_b+1}^{rn_b}\left(Z^{(1)}_{\boldsymbol{\alpha}_*}[j]+\log(1-\rho)\right)<\frac{b}{n_b}\right)\\
&\overset{(e)}{=}\prod_{r=1}^{\zeta}\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=(r-1)n_b+1}^{rn_b}Z^{(1)}_{\boldsymbol{\alpha}_*}[j]}{n_b}<I_*^{(1)}-\epsilon\right)
\end{aligned}\tag{20}$$

where (c) follows by binning the observations and bounding the maxima (see [5] and [16]), (d) follows from the independence of data across times conditioned on $\boldsymbol{S}$, and (e) follows from the definition of n_b. We then have that for any $b > 0$ from (19)

$$\sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > \zeta n_b)$$
$$\leq \sup_{S} \left[\prod_{r=1}^{\zeta} \mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=(r-1)n_b+1}^{rn_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} < I_*^{(1)} - \epsilon \right) \right]$$
$$= \left[\sup_{S} \mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} < I_*^{(1)} - \epsilon \right) \right]^{\zeta}. \tag{21}$$

By [28, Lemma 2], we have that

$$J \triangleq \mathbb{E}_{1,d}^{S} \left[\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} \right] = \frac{\sum_{j=1}^{n_b} \mathbb{E}_{p_{S^{(1)}[j]}} \left[Z_{\boldsymbol{\alpha}_*}^{(1)}[j] \right]}{n_b} \geq I_*^{(1)}. \tag{22}$$

This in turn implies that for any $\boldsymbol{S}$

$$\mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} < I_*^{(1)} - \epsilon \right)$$
$$= \mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} < I_*^{(1)} - \epsilon + J - J \right)$$
$$\leq \mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} < J - \epsilon \right)$$
$$\leq \mathbb{P}_{1,d}^{S} \left(\left| \frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} - J \right| > \epsilon \right). \tag{23}$$

Define

$$(\bar{\sigma}^{(1)})^2 \triangleq \max_{\boldsymbol{E} \in \mathcal{E}^{(1)}} \mathrm{Var}_{p_{\boldsymbol{E}}} \left[\log \frac{\overline{p}_{\boldsymbol{\alpha}_*^{(1)}}^{(1)}(\boldsymbol{X})}{g(\boldsymbol{X})} \right]. \tag{24}$$

From (14), we have that $(\bar{\sigma}^{(1)})^2 < \infty$. Then, by Chebychev's inequality

$$\mathbb{P}_{1,d}^{S} \left(\left| \frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} - J \right| > \epsilon \right) \leq \mathrm{Var}_{1,d}^{S} \left(\frac{\sum_{j=1}^{n_b} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{n_b} \right) \frac{1}{\epsilon^2}$$
$$= \frac{1}{\epsilon^2 n_b^2} \sum_{j=1}^{n_b} \mathrm{Var}_{p_{S^{(1)}[j]}} \left(Z_{\boldsymbol{\alpha}_*}^{(1)}[j] \right) \leq \frac{(\bar{\sigma}^{(1)})^2}{n_b \epsilon^2} \leq \delta \tag{25}$$

for large b, which, from (21), implies that for large b

$$\sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > \zeta n_b) \leq \delta^{\zeta}. \tag{26}$$

For the case of $\zeta > c_\epsilon$, we have that for large threshold b samples, $\boldsymbol{X}[1, \zeta n_b]$ can be generated in either phase 1 or phase 2. Define

$$t \triangleq \left\lceil \frac{I_*^{(1)}}{\min\{I_*^{(1)}, I_*^{(2)}\}} \right\rceil + 1. \tag{27}$$

We then have that for large b, $c_\epsilon n_b \leq \nu + d \leq (c_\epsilon + t) n_b$. Consider ζ such that $c_\epsilon + (m-1)t \leq \zeta \leq c_\epsilon + mt - 1$, for any $m \geq 1$. By following steps similar to (20)–(26), it can be established that for large b

$$\sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) \leq \delta^{c_\epsilon + m - 1}. \tag{28}$$

We then have from (19) and (26) that (28) and the definition of t

$$\sup_{S} \mathbb{E}_{1,d}^{S} \left[\frac{\tau_W(\boldsymbol{\alpha}_*, b)}{n_b} \right] \leq 1 + \sum_{\zeta=1}^{c_\epsilon} \delta^{\zeta} + \lim_{\xi \to \infty} \sum_{m=1}^{\xi} t \delta^{c_\epsilon + m - 1}$$
$$= \frac{1}{1-\delta} + t\delta^{c_\epsilon} + (t-1)\delta^{c_\epsilon + 1} \frac{1}{1-\delta} \triangleq 1 + \delta' \tag{29}$$

where $\delta' \to 0$ as $b \to \infty$ since $\delta \to 0$ as $b \to \infty$ and $c_\epsilon \geq 1$. This, in turn, implies that as $b \to \infty$

$$\mathrm{WADD}_d(\tau_W(\boldsymbol{\alpha}_*, b)) = \sup_{S} \mathbb{E}_{1,d}^{S} \left[\tau_W(\boldsymbol{\alpha}_*, b) \right]$$
$$\leq \frac{b}{I_*^{(1)}} (1 + o(1)). \tag{30}$$

Case 2: Consider the case of $c_1' \leq 1$. Define

$$n_b' \triangleq \left(d + \frac{b - \log \rho - d(I_*^{(1)} + \log(1-\rho))}{I_*^{(2)}} \right) (1 + \epsilon)$$
$$\sim \left(\frac{c_1'}{I_*^{(1)}} + \frac{1 - c_1'}{I_*^{(2)}} \right) (1 + \epsilon). \tag{31}$$

This implies that

$$\lim_{b \to \infty} \frac{n_b'}{d} = \left(1 + \left(\frac{1}{c_1'} - 1 \right) \frac{I_*^{(1)}}{I_*^{(2)}} \right) (1 + \epsilon) > 1$$

which, in turn, implies that for large b, $n_b' > d$ and $n_b' - d \to \infty$ as $b \to \infty$ [16]. By analyzing the expectation as in case 1, we then have that

$$\sup_{S} \mathbb{E}_{1,d}^{S} \left[\frac{\tau_W(\boldsymbol{\alpha}_*, b)}{n_b'} \right] \leq 1 + \sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > n_b')$$
$$+ \lim_{\xi \to \infty} \sum_{\zeta=2}^{\xi} \sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > \zeta n_b'). \tag{32}$$

Fix $\boldsymbol{S}$. Then since for any x, y and random variables X, Y, $\mathbb{P}(X + Y < x + y) \leq \mathbb{P}(X < x) + \mathbb{P}(Y < y)$, we have that

$$\mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > n_b')$$
$$= \mathbb{P}_{1,d}^{S} \Bigg(\sum_{j=1}^{d} \left(Z_{\boldsymbol{\alpha}_*}^{(1)}[j] + \log(1-\rho) \right) + \sum_{j=d+1}^{n_b'} Z_{\boldsymbol{\alpha}_*}^{(2)}[j]$$
$$< d(I_*^{(1)} + \log(1-\rho)) + (n_b' - d) I_*^{(2)} - \epsilon a \Bigg)$$
$$\leq \mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=1}^{d} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{d} < I_*^{(1)} - \frac{\epsilon a}{2d} \right)$$
$$+ \mathbb{P}_{1,d}^{S} \left(\frac{\sum_{j=d+1}^{n_b'} Z_{\boldsymbol{\alpha}_*}^{(2)}[j]}{n_b' - d} < I_*^{(2)} - \frac{\epsilon a}{2(n_b' - d)} \right)$$

where $a \triangleq dI_*^{(2)} + b - d(I_*^{(1)} + \log(1-\rho)) - \log\rho$. This, in turn, implies that

$$\begin{aligned}
&\sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > n_b) \\
&\leq \sup_{S} \mathbb{P}_{1,d}^{S}\left(\frac{\sum_{j=1}^{d} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{d} < I_*^{(1)} - \frac{\epsilon a}{2d}\right) \\
&+ \sup_{S} \mathbb{P}_{1,d}^{S}\left(\frac{\sum_{j=d+1}^{n_b'} Z_{\boldsymbol{\alpha}_*}^{(2)}[j]}{n_b' - d} < I_*^{(2)} - \frac{\epsilon a}{2(n_b' - d)}\right).
\end{aligned} \tag{33}$$

We now upper bound both of the terms in the right-hand side of (33). In particular, from (22), we have that

$$\begin{aligned}
&\sup_{S} \mathbb{P}_{1,d}^{S}\left(\frac{\sum_{j=1}^{d} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{d} < I_*^{(1)} - \frac{\epsilon a}{2d}\right) \\
&\leq \sup_{S} \mathbb{P}_{1,d}^{S}\left(\left|\frac{\sum_{j=1}^{d} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{d} - J\right| > \frac{\epsilon a}{2d}\right).
\end{aligned}$$

From Chebychev's inequality, we then have that

$$\begin{aligned}
&\sup_{S} \mathbb{P}_{1,d}^{S}\left(\left|\frac{\sum_{j=1}^{d} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]}{d} - J\right| > \frac{\epsilon a}{2d}\right) \\
&\leq \sup_{S} \frac{1}{d^2} \mathrm{Var}_{1,d}^{S}\left(\sum_{j=1}^{d} Z_{\boldsymbol{\alpha}_*}^{(1)}[j]\right)\left(\frac{2d}{\epsilon a}\right)^2 \\
&\leq \frac{1}{d}\left(\frac{2d\bar{\sigma}^{(1)}}{\epsilon a}\right)^2 \leq \frac{\delta}{2}
\end{aligned} \tag{34}$$

for large b since d/a converges to a constant and $d \to \infty$ as $b \to \infty$. Similarly, it can be shown that

$$\sup_{S} \mathbb{P}_{1,d}^{S}\left(\frac{\sum_{j=d+1}^{n_b'} Z_{\boldsymbol{\alpha}_*}^{(2)}[j]}{n_b' - d} < I_*^{(2)} - \frac{\epsilon a}{2(n_b' - d)}\right) \leq \frac{\delta}{2} \tag{35}$$

for large b.

Define

$$t' \triangleq \left\lceil \frac{1}{\left(\frac{c_1'}{I_*^{(1)}} + \frac{1-c_1'}{I_*^{(2)}}\right)\min\{I_*^{(1)}, I_*^{(2)}\}} \right\rceil + 1.$$

Following arguments similar to (27)–(30), we can establish that if $(m-1)t + 1 \leq \zeta \leq mt$ for any $m \geq 1$, we then have that

$$\sup_{S} \mathbb{P}_{1,d}^{S}(\tau_W(\boldsymbol{\alpha}_*, b) > \zeta n_b') \leq t'\delta^m. \tag{36}$$

Combining (32), (34), (35), and (36), we have that

$$\begin{aligned}
&\sup_{S} \mathbb{E}_{1,d}^{S}\left[\frac{\tau_W(\boldsymbol{\alpha}_*, b)}{n_b}\right] \leq 1 + \delta + \lim_{\xi\to\infty}\sum_{\zeta=2}^{\xi} t'\delta^{z-1} \\
&= \frac{1}{1-\delta} + t'\delta + (t'-1)\frac{\delta^2}{1-\delta} \triangleq \delta''
\end{aligned} \tag{37}$$

where $\delta'' \to 0$ as $b \to \infty$. As a result, we have that from (31) and (37)

$$\begin{aligned}
\mathrm{WADD}_d(\tau_W(\boldsymbol{\alpha}_*, b)) &= \sup_{S} \mathbb{E}_{1,d}^{S}\left[\tau_W(\boldsymbol{\alpha}_*, b)\right] \\
&\leq b\left(\frac{c_1'}{I_*^{(1)}} + \frac{1-c_1'}{I_*^{(2)}}\right)(1 + o(1)).
\end{aligned} \tag{38}$$

Finally, from (30) and (38), the theorem is established.

ACKNOWLEDGMENT

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] 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.*, Barcelona, Spain, 2020, pp. 8926–8930.

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

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

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

[5] 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.

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

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

[8] A. G. Tartakovsky and V. V. Veeravalli
Change-point detection in multichannel and distributed systems
Appl. Sequential Methodol.: Real-World Examples Data Anal., vol. 173, pp. 339–370, Jan. 2004.

[9] 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.

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

[11] Y. Mei
Quickest detection in censoring sensor networks
in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, 2011, pp. 2148–2152.

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

[13] 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.

[14] 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, Jun. 2016.

[15] G. Rovatsos, X. Jiang, A. D. Domínguez-García, and V. V. Veeravalli
Comparison of statistical algorithms for power system line outage detection
in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Shanghai, China, 2016, pp. 2946–2950.

[16] S. Zou, G. Fellouris, and V. V. Veeravalli
Quickest change detection under transient dynamics: Theory and asymptotic analysis
IEEE Trans. Inform. Theory, vol. 65, no. 3, pp. 1397–1412, Oct. 2018.

[17] 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, Jul. 2009.

[18] 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, Apr. 2010.

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

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

[21] 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, Mar. 2018.

[22] G. V. Moustakides and V. V. Veeravalli
Sequentially detecting transitory changes
in *Proc. IEEE Int. Symp. Inform. Theory*, Barcelona, Spain, 2016, pp. 26–30.

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

[24] 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*, Paris, France, 2019, pp. 2399–2403.

[25] 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.

[26] 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, 2019, pp. 98–102.

[27] 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. Inform. Theory*, Los Angeles, CA, USA, 2020, pp. 1171–1176.

[28] G. Rovatsos, G. V. Moustakides, and V. V. Veeravalli
Quickest detection of moving anomalies in sensor networks
, *IEEE J. Selected Areas Inf. Theory*, vol. 2, no. 2, pp. 762–773, Jun. 2021.

[29] A. Kott, A. Swami, and B. J. West
The internet of battle things
IEEE Comput., vol. 49, no. 12, pp. 70–75, Nov. 2016.

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

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

[32] V. V. Veeravalli and T. Banerjee
Quickest change detection
in *Proc. Academic Press Library Signal Process.*, Elsevier, 2014, pp. 209–255.

[33] T. M. Cover and J. A. Thomas
Elements of Information Theory. New York, USA: Wiley-Interscience, Jul. 2006.

[34] G. Lorden and M. Pollak
Sequential change-point detection procedures that are nearly optimal and computationally simple
Sequential Anal., vol. 27, no. 4, pp. 476–512, 2008.

[35] T. S. Lau, W. P. Tay, and V. V. Veeravalli
A binning approach to quickest change detection with unknown post-change distribution
IEEE Trans. Signal Proc., vol. 67, no. 3, pp. 609–621, Feb. 2019.

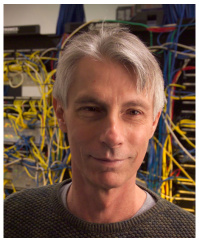
Georgios Rovatsos (Student Member, IEEE) received the B.S. degree from the University of Patras, Patras, Greece, in 2014, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2016, and 2020, respectively, all in electrical engineering.

His research interests are in statistical inference and machine learning.

Venugopal V. Veeravalli (Fellow, IEEE) received the B.Tech. degree (Silver Medal Honors) from the Indian Institute of Technology, Bombay, India, in 1985, the M.S. degree from Carnegie Mellon University, Pittsburgh, PA, USA, in 1987, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1992, all in electrical engineering.

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 the Coordinated Science Laboratory. Prior to joining the University of Illinois, he was on faculty of the ECE Department, Cornell University, Ithaca, NY, USA. He served as a Program Director for communications research with the U.S. National Science Foundation, Alexandria, VI, USA, 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.

Dr. Veeravalli was a Distinguished Lecturer for the IEEE Signal Processing Society from 2010 to 2011. He has been on the Board of Governors of the IEEE Information Theory Society. He has been an Associate Editor for *Detection and Estimation*, IEEE TRANSACTIONS ON INFORMATION THEORY, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. 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 (PECASE), and the Wald Prize in Sequential Analysis.



Don Towsley (Life Fellow, IEEE) received the B.A. degree in physics and the Ph.D. degree in computer science from University of Texas, Austin, TX, USA, in 1971 and 1975, respectively.

He is currently a Distinguished Professor with the College of Information & Computer Sciences, University of Massachusetts, Amherst, MA, USA. He has held visiting positions with numerous universities and research labs. His research interests include network science, performance evaluation, and quantum networking.

He is Cofounder and Coeditor-in-Chief for the new *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* (TOMPECS), and has served as Editor-in-Chief for *IEEE/ACM Transactions on Networking* and on numerous editorial boards. He has served as Program Cochair of several conferences including *INFOCOM 2009*. He is a corresponding member of the Brazilian Academy of Sciences and is the recipient of several achievement awards including the 2007 IEEE Koji Kobayashi Award and the 2011 INFOCOM Achievement Award. He is the recipient of numerous paper awards including the 2012 ACM SIGMETRICS Test-of-Time Award, a 2008 SIGCOMM Test-of-Time Paper Award, and a 2018 SIGMOBILE Test-of-time Award. He is also a recipient of the 1998 IEEE Communications Society William Bennett Best Paper Award. Last, he has been elected as Fellow of the ACM.

Ananthram Swami (Life Fellow, IEEE) received the B.Tech. degree from IIT-Bombay, Bombay, India, the M.S. degree from Rice University, Houston, TX, USA, and the Ph.D. degree from the University of Southern California (USC), Los Angeles, Ca, USA, all in electrical engineering.

He is with the U.S. Army Research Laboratory (ARL), Adelphi, MD, USA, and is the Army's Senior Research Scientist (ST) for Network Science. Prior to joining ARL, he held positions with Unocal Corporation, USC, and CS-3. He has held visiting faculty positions with INP, Toulouse, France, and with Imperial College, London, U.K.

Dr. Swami is a recipient of the 2018 IEEE ComSoc MILCOM Technical Achievement Award and the 2017 Presidential Rank Award (Meritorious). He is an ARL Fellow.