DYNAMIC ANOMALY DETECTION IN SENSOR NETWORKS

BY

GEORGIOS ROVATSOS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Venugopal V. Veeravalli, Chair
Professor Minh N. Do
Professor Alejandro D. Domínguez-García
Associate Professor Georgios Fellouris

ABSTRACT

In the problem of quickest change detection, a sequence of random variables is observed sequentially by a decision maker. At some unknown time instant, the emergence of an anomaly leads to a change in the distribution of the observations. The goal in quickest change detection is to detect this change as quickly as possible, subject to constraints on the frequency of false alarm events. One important application of the theory of quickest change detection is in the context of anomaly detection in sensor networks used to monitor engineering systems. Sensor network related detection problems can vary significantly depending on the spatial evolution of the anomaly in the network as time progresses. Settings involving static anomalies, i.e., anomalies that are perceived by all sensors concurrently and that affect sensors persistently, have been extensively studied in the quickest change detection literature. In addition, semi-dynamic quickest change detection settings that involve anomalies that affect sensors at different time instants, albeit in a persistent manner, have recently received more attention. In this dissertation, our goal is to study the problem of dynamic anomaly detection in sensor networks, i.e., the case where anomalies may not affect sensors persistently, but may move around the network affecting different sets of sensors with time. The objective is to design anomaly detection procedures that are provably optimal with respect to delay-false alarm trade-off formulations. We study the quickest dynamic anomaly detection problem under multiple settings by imposing different assumptions on the spatial evolution of the anomaly. In particular, we consider the case where anomalies evolve according to a discrete-time Markov chain model, for which we develop asymptotically optimal procedures which we compare with more computationally feasible heuristic detection algorithms that require less model knowledge. The Markov model definition incorporates anomalies the size of which may be constant or vary with time. In addition, we study the worst-path dynamic anomaly detection setting, where we assume that the trajectory of the anomaly is unknown and deterministic, and that candidate detection procedures are evaluated according to the anomaly path that maximizes their detection delay. We consider the worst-path setting under the assumption that the anomaly affects a fixed size of sensors, as well as study the problem of worst-path anomaly detection when the size of the anomaly changes with time. For the two worst-path settings we establish that algorithms from quickest change detection literature can be modified to result in provably asymptotically optimal, and in some cases, exactly optimal procedures. A detailed performance analysis of the proposed algorithms is conducted, and concise guidelines regarding the design of proposed tests are provided. Numerical studies of the proposed detection schemes are presented for all studied settings and for a variety of test cases, such as different network sizes, probability distributions, and degrees of model knowledge. Finally, we outline problems of interest for future work, such as the extension of proposed algorithms and techniques in settings where model knowledge is limited.

To my mother, for her endless love and support.

ACKNOWLEDGMENTS

I would like to thank Prof. Veeravalli for advising me during my studies at UIUC and for our wonderful collaboration. I am very grateful for the time and effort he devoted to help me become a better researcher. His mentoring was pivotal in helping me understand the responsibilities of a researcher, both towards the academic community as well as themselves, and in shaping my understanding of how to work in order to improve myself on a daily basis. I am also very grateful to Prof. Moustakides for our collaboration during my doctoral studies. His advice and mentoring played an enormous role in the shaping of this dissertation. Furthermore, I would like to thank him for steering my interest towards the area of probability and statistics during my undergrad studies. I am very thankful to Prof. Do, Prof. Domínguez-García, and Prof. Fellouris for agreeing to participate in my doctoral committee and for taking the time to read my dissertation and provide valuable feedback.

I would like to thank Prof. Zou, Prof. Domínguez-García, Prof. Yiang, and Prof. Towsley for our collaboration on a variety of research problems during my years at UIUC. Furthermore, I am grateful to Prof. Fellouris for his suggestions and for helpful discussions regarding my Ph.D., as well as my M.Sc. research work.

I would like to thank the UIUC, and especially the Department of Electrical and Computer Engineering, for giving me the opportunity to learn so much by taking courses and by interacting with some of the best researchers in the world. I am grateful to the Coordinated Science Laboratory for providing an outstanding research environment. The faculty affiliated with CSL and the ECE department were always eager to provide valuable advice without hesitation, something which made my experience working as a graduate student unique. Furthermore, I owe my gratitude to the staff of CSL for always being there to help me during my studies. I would also like to thank my CSL colleagues, and especially the members of Prof. Veeravalli's research group, for creating a collaborative environment where all students could work together and become better researchers.

I my very thankful to all the friends I made during my time in Urbana-Champaign for spending time together and for motivating me during my graduate studies. Meeting and bonding with people from so many different ethnic as well as academic backgrounds was definitely one of the most important aspects of my studies, and I would like to thank the UIUC for providing such a diverse environment. Furthermore, I am very grateful to my friends from Greece for supporting me so much during my graduate years, and for their friendship throughout the years.

Finally, I would like to thank my family for their unconditional support and for always believing in me. I am especially grateful to my mother for instilling in me the understanding of the importance of a good education.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In many engineering applications, maintaining an accurate estimate of the state of the monitored system is crucial to ensure reliable operation. In such settings, the goal is often to detect whether an anomaly has led the system to enter an abnormal state. Engineering applications where such real-time decision making is crucial range from the detection of subtle faults that may lead to catastrophic failures in large-scale systems to applications in financial surveillance [1–9]. In modern engineering systems, state inference is frequently facilitated by sequentially obtaining measurements from fast sampling units monitoring the system. The scale of the data obtained due to these fast sampling operations and the growing size of engineering systems render the algorithmic detection of anomalies necessary. In statistics, the design of sequential detection algorithms is frequently studied within the framework of *quickest change detection* (QCD) [10–12].

The goal in QCD is to detect a change in the distribution of a sequentially observed process as quickly as possible, subject to a tolerable *false alarm* (FA) constraint. This change happens at an unknown time instant, referred to as the *changepoint*. To our knowledge, the earliest instances of QCD algorithms were the *Shewhart* test [13] and the celebrated *Cumulative-Sum* (CUSUM) test [14]. These tests were not proposed with the goal of designing procedures that are provably optimal, but mostly as heuristic schemes to be used in monitoring manufacturing processes.

In the classical QCD problem ([10–12]), the statistical behavior of the observed process is completely specified by the *non-anomalous* and *anomalous* distributions, that generate *independent and identically distributed* (i.i.d.) observations before and after the emergence of the anomaly in the system. This classical QCD setting is often referred to as the *i.i.d. model*. The first major theoretical study of QCD algorithms was conducted by Shiryaev in [15, 16]. In these works, Shiryaev introduced the *Bayesian* version of the classical *single-changepoint* QCD problem, where the changepoint is modeled as a random variable with a known distribution and the goal is to minimize the average detection delay subject to a bound on the probability of FA. He then proposed the *Shiryaev* test as the exact solution to the Bayesian problem under the i.i.d. model assumption. Later, Lorden introduced the *minimax* QCD setting where the changepoint is considered to be an *unknown* but *deterministic* quantity, and the goal is to minimize a *worst-case average detection delay* (WADD) subject to a lower bound on the *mean time to false alarm* (MTFA) [17]. Lorden established that under the i.i.d. assumption Page's CUSUM test is asymptotically optimal with respect to the minimax constrained optimization problem as the MTFA goes to infinity. Pollak proposed using a less pessimistic detection delay metric for the minimax QCD setting, and established that the Shiryaev procedure can be modified to solve the resulting delay-FA trade-off asymptotically in the i.i.d. setting [18]. The first non-asymptotic result for minimax QCD was provided by Moustakides in [19], where it was shown that the CUSUM test is the exact solution to Lorden's optimization problem. Moustakides later also established that the CUSUM procedure is exactly optimal for a special case of dependent processes [20]. Poor showed that the CUSUM algorithm is also optimal when the detection delay is penalized exponentially rather than linearly, as was the case in previous formulations of the classical i.i.d. setting [21]. Lai developed a unified approach to Lorden's and Pollak's delay-FA problem formulations and established that the CUSUM test is asymptotically optimal for both the optimization problems even when the observations are dependent [22]. A Bayesian version of the single-sensor QCD problem for the case of non-i.i.d. data was proposed by Tartakovsky and Veeravalli in [23], where it was shown that the Shiryaev test can be modified to provide a solution that is asymptotically optimal as the probability of false alarm goes to zero, even when the distribution of the changepoint is not necessarily geometric.

The earliest theoretical results in QCD were derived for the single-sensor setting, i.e., when the observations are sampled by one sensor that is eventually affected by a change in distribution. In this dissertation, the objective is to study the problem of anomaly detection in *sensor networks*, i.e., groups of sensors used by a decision maker to monitor an engineering system in real-time. Sequential detection problems in the context of sensor networks have been extensively studied in the literature. Detailed research has been conducted when the monitored engineering system is affected by a *static* anomaly, i.e., an anomaly that is perceived simultaneously and that leads to a persistent change in the data-generating distributions of the affected sensors. In particular, in [6] it was established that running a CUSUM algorithm at each node and declaring a change as soon as an alarm is raised by any of the sensors provides an asymptotically optimal procedure for the case of a single sensor being affected by the anomaly. In [24], Mei proved that for the case of i.i.d. observations before and after the changepoint an asymptotically optimal procedure can be derived even when the anomaly affects an unknown subset of nodes of unknown size persistently. His procedure was based on calculating a CUSUM statistic at each node and comparing the sum of the node statistics to a threshold to decide whether to raise an alarm. For the same setting an alternative procedure that was shown to perform better than Mei's SUM-CUSUM scheme was proposed in [25]. In [26], the aforemetioned setting of an unknown set of multiple affected sensors was studied in a more general framework and second-order asymptotically optimal algorithms were proposed. In [27], it was established that when an upper bound on the number of affected sensors is known to the decision maker, Mei's SUM-CUSUM scheme can be modified to result in an asymptotically optimal test.

In addition to static anomaly detection, the problem of sequentially detecting *semi-dynamic* anomalies, i.e., anomalies that affect sensors persistently but may be perceived at different time instants across sensors, has recently received attention in the literature [28–34]. An important work in this subject is [32], where it was established that Mei's SUM-CUSUM test is asymptotically optimal under i.i.d. assumptions even when the affected sensors do not perceive the anomaly simultaneously. In addition, the semi-dynamic anomaly detection setting was also considered in [31], where the authors designed asymptotically optimal procedures to detect anomalies after they have affected more than a pre-determined number of sensors, and not directly after their emergence as in [32]. Distributed versions of the algorithms in [31] were introduced in [33, 34].

Note that all the aforementioned sensor network problems have a common element: there is a *persistent* change in the distribution of each affected sensor after it perceives the anomaly, even if the anomaly does not affect sensors concurrently. However, the problem of detecting anomalies that are *dynamic* in nature has not been studied in the literature. The crucial difference between the current literature on sensor network event detection and our proposed dynamic anomaly setting is that in our work the anomaly need not be persistent in any specific node, but it is persistent if we view the entire network as a whole. This means that the anomaly is moving, implying that the sets of affected sensors may vary as time progresses, and each sensor can shift between the non-anomalous and anomalous state (we will refer to such anomalies by using the term *dynamic* or *moving* anomaly interchangeably).

There is a plethora of applications that we believe the dynamic anomaly setting fits to, ranging from video object detection to detection of physically moving adversaries. The formulations presented in this dissertation are particularly relevant to adversarial settings, since in practice adversaries may attempt to mask the emergence of an anomaly by forcing it to affect different parts of the network as time progresses. Our goal in this work is to formulate a family of dynamic anomaly detection problems and propose solutions that are tractable, as well theoretically justified with respect to QCD trade-off formulations of practical and theoretical interest. In particular, we study three different settings: i) the case of a dynamic anomaly that evolves according to a *discrete-time Markov chain* (DTMC); ii) the case of a dynamic anomaly of constant size where there is no prior statistical knowledge concerning the trajectory of the anomaly; iii) the case of a dynamic anomaly that varies in size, affecting a different number of sensors as time progresses. In particular, the dissertation is organized as follows:

1. In Chapter 2, we study the problem of sequentially detecting dynamic anomalies under a Markov evolution assumption. In particular, we begin introducing the Markov anomaly model, including the main requirements that it has to satisfy to proceed with the analysis. We then frame the underlying QCD problem as a dynamic composite hypothesis testing problem and construct a *Windowed Generalized Likelihood Ratio* (Windowed-GLR) test to detect the emergence of the anomaly. We establish that the proposed test is asymptotically optimal with respect to a defined delay-FA framework. We then compare its performance with other asymptotically optimal and heuristic procedures. We conclude that a CUSUM-type procedure, that may not

necessarily be asymptotically optimal, offers very good performance in comparison to provably asymptotically optimal procedures while being more computationally efficient and requiring significantly less model knowledge. It should be noted that procedures developed in this chapter can handle both the cases of anomalies of constant and varying size.

2. The Markov chain setting discussed in Chapter 2 can be non-practical in engineering applications since proposed procedures require complete knowledge of the Markov transition rates, something which might not be feasible, especially for the case of large sensor networks. In Chapter 3, we lift the Markov anomaly evolution assumption and consider the problem of *worst-path* anomaly detection for anomalies of constant size. We begin by introducing a worst-path modification of Lorden's detection delay metric [17] used in Chapter 2. According to the modified metric, candidate detection schemes are evaluated with respect to the anomaly trajectory that maximizes their detection delay. We then establish that the CUSUM test introduced in Chapter 2 is an exact solution to the studied problem for the case of a homogeneous sensor network when the parameters of test are chosen to be equal. In addition, we prove that for the case of heterogeneous sensors the parameters of the proposed CUSUM test can be carefully chosen to yield a first-order asymptotically optimal test. An interesting observation is that the resulting algorithm is an equalizer rule with respect to the placement of the anomaly when considering the asymptotic performance of the proposed CUSUM procedure. We conclude this chapter by comparing the proposed test with heuristic as well as oracle algorithms that require complete knowledge of the trajectory of the anomaly. Furthermore, we numerically investigate the performance loss that our test incurs when the weights are not chosen optimally.

3. In Chapter 4, we consider the problem of sequential detection of worst-path *varying-size* dynamic anomalies, i.e., anomalies that move around the network while affecting a different number of nodes as time progresses. In particular, we study the setting where the anomaly does not settle to a persistent anomaly size instantly, but after a series of transient phases. Each transient phase corresponds to a different

anomaly size and each one starts from a respective changepoint. We begin the chapter by presenting the observation model. We then introduce a generalization of the delay metric used in Chapter 3 that takes the presence of transient phases and their durations into consideration. We establish that a *Weighted Dynamic* CUSUM-type (WD-CUSUM) test [35] is asymptotically optimal for a specific choice of algorithm parameters. This choice leads to a procedure that is an equalizer rule with respect to anomaly placement at each post-change phase, a conclusion similar to the one obtained for the test studied in Chapter 3. We conclude by numerically evaluating the performance of our proposed algorithm for different network sizes and for varying degrees of model knowledge.

A QCD problem related to our work, especially compared to the dynamic anomaly detection problem presented in Chapter 2, is the problem of QCD in *hidden Markov models* (HMMs). Although the HMM QCD setting was not initially studied in the context of sensor networks, algorithms and analytical techniques can be exploited and used in the setting of Markov anomaly detection. HMM QCD has been studied in prior work, e.g., see [36–40]. In [36–38], the problem of minimax HMM QCD was studied. For this problem, the GLR-based test does not have a recursion, and is thus not computationally efficient. In [36], instead of using the GLR approach, a recursive test was designed using an approximate conditional probability distribution, and was further shown to be first-order asymptotically optimal. The main differences between our Markov anomaly setting and the work in [36–38] are the following: (i) we focus on the application of sequential dynamic anomaly detection in sensor networks; (ii) the work in [36–38] considers the setting where the observations are generated according to a HMM, and at some unknown but deterministic time, the parameters of the HMM change abruptly, whereas in our problem, the data before the changepoint is i.i.d. distributed, the data after the change is generated by a HMM, and the pre-change data is independent from the post-change data; (iii) we construct a Windowed-GLR test and establish its first-order asymptotic optimality using a technique introduced in [22]; (iv) we also construct several alternative algorithms, including the *Dynamic Shiryav-Roberts* (D-S-R) test, the QCD test with changepoint estimation, and the *Mixture-CUSUM* (M-CUSUM) algorithm; and (v) we comprehensively compare these algorithms numerically, and investigate the conditions under which each of these tests should be preferred. The recursive HMM test of [36] was further studied in [38] for two-state HMMs, where it was shown to be equivalent to a quasi-GLR scheme with respect to a pseudo post-change measure. Recently, the Bayesian setting was investigated in [39], where the changepoint is modeled as a random variable with known distribution. Our analysis for the case of an anomaly that evolves according to a DTMC mainly uses the theoretical results of [22] and [39]. Although the work in [39] focuses on the Bayesian case, some of the convergence results provided can still be employed in our minimax setting. A different formulation of QCD in HMMs was proposed in [40], and Shewhart-type tests were constructed and were shown to exactly maximize the worst-case detection probability subject to false alarm constraints.

Furthermore, our work is related to the single-sensor QCD problem under transient dynamics studied in [2, 35, 41, 42], where the change in the probability distribution of the observations does not happen instantaneously, but through a sequence of transient phases each corresponding to a distinct data-generating distribution. In particular, as will be seen in Chapter 4, the varying-size dynamic anomaly detection problem involves transient phases in the sense of [35]. However, note that in the dynamic anomaly setting the statistical behavior of the observed process during these transient phases is not completely specified, since the location of the anomalous nodes is not known by the decision maker. As a result, it is not apparent whether the detection procedures in [35] can be directly applied to the studied dynamic anomaly detection setting. Furthermore, in our work, we do not make the assumption that there is a persistent statistical behavior after a specific time instant, unlike [35] where it is assumed that the system reaches a persistent phase during which the distribution of the data does not change. Our main assumption is that the anomaly eventually settles to a persistent anomaly size. Although the varying-size anomaly detection and the transient QCD settings appear to have significant differences, in Chapter 4 we establish that a solution to a specific instance of the transient QCD problem presented in [35] is also an asymptotic solution to the varying-size dynamic anomaly detection problem.

Notation

Before proceeding to the main part of the dissertation, we introduce some necessary notation. A main assumption in this work is that all sensor observations take values in $\mathbb{R}$. To this end, let $\mathcal{B}(\mathbb{R}^L)$ denote the Borel σ-algebra with respect to $\mathbb{R}^L$, $L \geq 1$, and let μ a σ-finite measure on $\mathbb{R}^L$. Our convention in this work is that 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, 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 $[K]$. The sequence $\boldsymbol{X} \triangleq \{\boldsymbol{X}[k]\}_{k=1}^{\infty}$ denotes the sequence of random variables 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 ℓ at time k. Furthermore, $\sigma(\boldsymbol{X}[k_1, k_2])$ denotes the σ-algebra generated by $\boldsymbol{X}[k_1, k_2]$. Define the Gaussian distribution with mean θ and variance σ^2 by $\mathcal{N}(\theta, \sigma^2)$. Denote by $D(f\|g)$ the Kullback-Leibler divergence [43] between two probability density functions $f(\cdot)$ and $g(\cdot)$. Furthermore, for $K \geq 0$, $\|\boldsymbol{x}\|_K$ denotes the K-norm of vector $\boldsymbol{x}$. $(x)^+$ restricts x from taking negative values, i.e., $(x)^+ \triangleq \max\{x, 0\}$. Finally, for functions $f : \mathbb{R} \mapsto \mathbb{R}$, $g : \mathbb{R} \mapsto \mathbb{R}$, $f(x) \sim g(x)$ denotes that $\frac{f(x)}{g(x)} = 1 + o(1)$ as $x \to \infty$, where $o(1) \to 0$ as $x \to \infty$.

CHAPTER 2

MARKOV DYNAMIC ANOMALY DETECTION

In this chapter, we consider the problem of sequential detection of a dynamic anomaly that evolves according to a discrete-time Markov chain. In this setting, the Markov transition rates quantify the probability of moving from a specific set of anomalous nodes to a different set during the post-change regime. We begin by introducing the observation model that describes the data generation process. Following, we introduce the Markov model that governs the evolution of the anomaly, along with assumptions the model needs to satisfy so that our analysis is valid. We proceed by presenting the delay-FA optimization problem to be solved. In this chapter, we consider both Lorden's and Pollak's delay metrics, under the understanding that these metrics will depend on the underlying Markov model. We then frame the studied QCD problems as a dynamic composite hypothesis testing problem and introduce the Windowed-GLR test, which we establish to be first-order asymptotically optimal under both Lorden's and Pollak's formulations. In addition, we present other algorithms that vary in terms of algorithmic performance and in terms of model knowledge they require, and compare them to the Windowed-GLR test. Numerical results imply that a recursive CUSUM-type procedure offers comparable performance to provably asymptotically optimal procedures while being more computationally efficient and requiring significantly less model knowledge. This chapter has appeared in part as [44, 45].

2.1 Observation Model

Consider a network of $L \geq 1$ nodes denoted by $[L] \triangleq \{1, \ldots, L\}$. Denote by $g_\ell(x)$, $f_\ell(x)$ the *non-anomalous* and *anomalous probability density functions* (pdfs) at sensor $\ell \in [L]$, respectively. We assume that at each sensor the corresponding non-anomalous and anomalous 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 non-anomalous distribution, and observations are assumed to be independent across sensors. As a result, the joint pdf of $\boldsymbol{X}[k]$ is initially given by

$$g(\boldsymbol{X}[k]) \triangleq \prod_{\ell=1}^{L} g_\ell(X_\ell[k]). \tag{2.1}$$

After some *unknown* and *deterministic changepoint* $\nu \geq 1$, a physical event leads to the emergence of a dynamic anomaly in the network. The anomaly moves around the network, affecting different sets of size $m \in [L]$ 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 $k \geq \nu$. We denote by $\mathcal{E} \triangleq \left\{\boldsymbol{E}_j \,\middle|\, 1 \leq j \leq \binom{L}{m}\right\}$ 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). Nodes affected by the anomaly generate observations according to the anomalous pdf. In particular, for $k \geq \nu$, we have that *conditioned* on $\boldsymbol{S}$, the joint pdf of $\boldsymbol{X}[k]$ is given by

$$p_{\boldsymbol{S}[k]}(\boldsymbol{X}[k]) \triangleq \left(\prod_{\ell \in \boldsymbol{S}[k]} f_\ell(X_\ell[k])\right) \cdot \left(\prod_{\ell \notin \boldsymbol{S}[k]} g_\ell(X_\ell[k])\right), \tag{2.2}$$

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 assume that the observations before the changepoint are independent from the observations after the changepoint. However, whether the data are independent across time after the changepoint depends on our assumption on $\boldsymbol{S}$. For example, as will be seen in this chapter, assuming that $\boldsymbol{S}$ evolves according to a

Markov model implies that the observations after the changepoint will not be independent across time. According to eqs. (2.1) and (2.2), *conditioned* on ν and $\boldsymbol{S}$ the complete statistical model is then the following:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]), & 1 \leq k < \nu, \\ p_{\boldsymbol{S}[k]}(\boldsymbol{X}[k]), & k \geq \nu. \end{cases} \tag{2.3}$$

2.2 Markov Trajectory Model

In this chapter, we study the setting where the evolution of $\boldsymbol{S}[k]$ is specified by a *discrete-time Markov chain* (DTMC). More specifically, we denote by $\mathbb{P}_\nu(\cdot)\,(\mathbb{E}_\nu[\cdot])$ the probability measure (expectation) when the anomaly occurs at time ν. In addition, we denote by $\mathbb{P}_\infty(\cdot)\,(\mathbb{E}_\infty[\cdot])$ the probability measure (expectation) when $\nu = \infty$, i.e., when there is no anomaly. Then, for any $k \geq \nu$, under the Markov assumption we have that

$$\mathbb{P}_\nu(\boldsymbol{S}[k+1]|\boldsymbol{S}[1,k], \boldsymbol{X}[1,k]) = \mathbb{P}_\nu(\boldsymbol{S}[k+1]|\boldsymbol{S}[k]) \triangleq \lambda_{\boldsymbol{S}[k],\boldsymbol{S}[k+1]}, \tag{2.4}$$

where $\lambda_{\boldsymbol{E},\boldsymbol{E}'} \in [0,1]$ denotes the probability that the anomaly placement changes from the one in $\boldsymbol{E}$ to the one in $\boldsymbol{E}'$ for $\boldsymbol{E}, \boldsymbol{E}' \in \mathcal{E}$. Furthermore, for any $k \geq \nu$, conditioned on $\boldsymbol{S}[k]$, $\boldsymbol{X}[k]$ is independent from anything else. To be more explicit, for any $\boldsymbol{B} \in \mathcal{B}(\mathbb{R}^L)$, we have that

$$\mathbb{P}_\nu(\boldsymbol{X}[k] \in \boldsymbol{B}|\boldsymbol{X}, \boldsymbol{S}) = \mathbb{P}_\nu(\boldsymbol{X}[k] \in \boldsymbol{B}|\boldsymbol{S}[k]) = \int_B p_{\boldsymbol{S}[k]}(\boldsymbol{x}) d\mu(\boldsymbol{x}). \tag{2.5}$$

Under the Markov evolution assumption, the underlying stochastic process of this problem can be viewed as a *hidden Markov model* (HMM), where $\{\boldsymbol{S}[k]\}_{k=\nu}^\infty$ is a finite-state Markov chain, which is not directly observable. The transition probability matrix is given by $[\lambda_{\boldsymbol{E},\boldsymbol{E}'}]_{\boldsymbol{E},\boldsymbol{E}' \in \mathcal{E}}$. Then, the sequence of random vectors $\{\boldsymbol{X}[k]\}_{k=\nu}^\infty$ is adjoint to this Markov chain according to (2.4) and (2.5). Therefore, after the anomaly appears in the network, there is a change in the underlying stochastic process from an i.i.d. model to a HMM.

In order to proceed with our analysis, we make the following assumptions on the DTMC. In particular we assume that:

(C.1) Under $\mathbb{P}_1(\cdot)$, the DTMC $\boldsymbol{S}$ is *ergodic* (positive recurrent, irreducible and aperiodic) [46]. Furthermore, if we define the random matrices

$$\mathsf{M} \triangleq \operatorname{diag}\left(\{p_{\boldsymbol{E}}(\boldsymbol{X}[1]) : \boldsymbol{E} \in \mathcal{E}\}\right) \tag{2.6}$$

and

$$\mathsf{N} \triangleq \left[\lambda_{\boldsymbol{E},\boldsymbol{E}'} p_{\boldsymbol{E}'}(\boldsymbol{X}[2])\right]_{\boldsymbol{E},\boldsymbol{E}' \in \mathcal{E}}, \tag{2.7}$$

then M and N are almost surely invertible under $\mathbb{P}_1(\cdot)$ and $\mathbb{P}_\infty(\cdot)$.

Note that according to C.1 the DTMC $\boldsymbol{S}$ defined in (2.4) has a stationary distribution denoted by a vector $\boldsymbol{\alpha} \triangleq \{\alpha_{\boldsymbol{E}} : \boldsymbol{E} \in \mathcal{E}\} \in \mathcal{A}$ [46]. Here, $\mathcal{A}$ denotes the simplex of all probability vectors of dimension $|\mathcal{E}|$. We also assume that $\boldsymbol{S}$ is initialized with $\boldsymbol{\alpha}$, i.e., that for all $\boldsymbol{E} \in \mathcal{E}$, $\mathbb{P}_\nu(\boldsymbol{S}[\nu] = \boldsymbol{E}) \triangleq \alpha_{\boldsymbol{E}}$.

(C.2) There exists $r > 0$ such that

$$\int_{\mathbb{R}} |x|^{r+1} g_\ell(x) dx < \infty, \tag{2.8}$$

and

$$\int_{\mathbb{R}} |x|^{r+1} f_\ell(x) dx < \infty, \tag{2.9}$$

for all $\ell \in [L]$. The above assumptions cover many interesting examples of HMMs, as noted in [39].

Note that from C.1 and the observation and Markov models in Sections 2.1 and 2.2, we have that for $k_1 \leq \nu \leq k_2$

$$\begin{aligned} &h_\nu(\boldsymbol{X}[k_1, k_2]) \\ &\triangleq g(\boldsymbol{X}[k_1, \nu-1]) \cdot \sum_{\boldsymbol{E}_\nu, \ldots, \boldsymbol{E}_{k_2} \in \mathcal{E}} \left\{ \alpha_{\boldsymbol{E}_\nu} p_{\boldsymbol{E}_\nu}(\boldsymbol{X}[\nu]) \cdot \prod_{j=\nu+1}^{k_2} \left[\lambda_{\boldsymbol{E}_{j-1},\boldsymbol{E}_j} p_{\boldsymbol{E}_j}(\boldsymbol{X}[j])\right] \right\} \end{aligned} \tag{2.10}$$

denotes the joint probability distribution of $\boldsymbol{X}[k_1, k_2]$ conditioned on a changepoint ν.

Remark 1. *Note that, although the observation model in this chapter was outlined for the case of dynamic anomalies of constant size, the current* *setting can be easily extended to anomalies of varying size easily by modifying the definition of the DTMC to include states of varying size and location. Although this extension to varying-size anomalies is straightforward for the Markov evolution setting, in this chapter in order to facilitate the presentation of the material we focus on the case of constant-size anomalies. However, all the algorithms presented in this chapter can be directly applied to the case of varying-size dynamic anomalies by modifying the definition of the underlying DTMC. As will be seen later in this dissertation, such an extension from the setting of constant-size to varying-size anomalies is not as straightforward for the worst-path setting studied in Chapters 3 and 4.*

2.3 Problem Formulation

The goal in QCD is to design stopping times that can detect the emergence of an anomaly as quickly as possible while ensuring that the frequency of FA events is below an acceptable level. In QCD, detection procedures take the form of *stopping times* [10–12]. A stopping time τ with respect to the observed sequence $\boldsymbol{X}$ is an integer-valued random variable, such that for each $k \geq 1$, $\{\tau \leq k\} \in \sigma(\boldsymbol{X}[1,k])$. In other words, the decision to stop at time k is determined only by $\boldsymbol{X}[1,k]$.

In this dissertation, we focus on *minimax* problem settings, where the changepoint ν is assumed to be *deterministic* and *unknown*. In order to measure the frequency of false alarm events, we define the *mean time to false alarm* (MTFA) as

$$\mathbb{E}_\infty[\tau]. \tag{2.11}$$

The definition of the detection delay is dependent on the way the trajectory of the anomaly is modeled. For the Markov case studied in this chapter, our detection delays will depend on the underlying Markov model. To this end, we begin by defining the *worst-case average detection delay* (WADD) of a stopping time τ under Lorden's criterion, introduced in [17], for the Markov model case by

$$\text{WADD}(\tau) \triangleq \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu\big[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\big], \tag{2.12}$$

where the convention that $\mathbb{E}_\nu\left[\tau-\nu+1|\tau\geq\nu,\boldsymbol{X}[1,\nu-1]\right]\triangleq 1$ when $\mathbb{P}_\nu(\tau\geq\nu)=0$ is used. In addition, define the *conditional average detection delay* (CADD) under Pollak's criterion (see [18]) by

$$\mathrm{CADD}(\tau)\triangleq\sup_{\nu\geq 1}\mathbb{E}_\nu\big[\tau-\nu|\tau\geq\nu\big]. \tag{2.13}$$

The WADD metric is a more pessimistic metric than the CADD metric (for more details see, e.g., [11]); in particular, it can be shown that for any stopping rule τ

$$\mathrm{WADD}(\tau)\geq\mathrm{CADD}(\tau). \tag{2.14}$$

In this dissertation, we aim to design stopping rules that minimize a given detection delay subject to a constraint on the MTFA. In particular, define the family of stopping times

$$\mathcal{C}_\gamma\triangleq\{\tau:\mathbb{E}_\infty[\tau]\geq\gamma\}, \tag{2.15}$$

i.e., stopping times that satisfy the MTFA constraint for a pre-determined constant $\gamma>0$. Our goal in this chapter is to design stopping rules that solve the following constrained optimization problems:

$$\begin{aligned}\min_{\tau}\quad & \mathrm{WADD}(\tau)\\ \text{s.t.}\quad & \tau\in\mathcal{C}_\gamma\end{aligned} \tag{2.16}$$

and

$$\begin{aligned}\min_{\tau}\quad & \mathrm{CADD}(\tau)\\ \text{s.t.}\quad & \tau\in\mathcal{C}_\gamma.\end{aligned} \tag{2.17}$$

Remark 2. *In this dissertation, 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.*

2.4 Windowed-GLR Test

In QCD, algorithms are frequently designed by framing the problems studied in a dynamic composite hypothesis testing setting and constructing a test based on the *generalized likelihood ratio* (GLR) statistic [10–12]. In particular, in the Markov setting studied in this chapter at each time k we distinguish between the following two hypotheses:

$$\mathscr{H}_0^k : \text{the anomaly appears at time } \nu > k, \tag{2.18}$$

$$\mathscr{H}_1^k : \text{the anomaly appears at time } \nu \leq k. \tag{2.19}$$

Note that under the alternative hypothesis the changepoint ν is unknown. We then take a GLR approach to construct the detection statistic (see, e.g., [11] for the interpretation of classical QCD tests through the GLR approach). Specifically, the likelihood under these two hypotheses can be expressed respectively as follows:

$$\mathscr{H}_0^k : \prod_{j=1}^{k} g(\boldsymbol{X}[j]), \tag{2.20}$$

$$\mathscr{H}_1^k : \prod_{j=1}^{\nu-1} g(\boldsymbol{X}[j]) \prod_{j=\nu}^{k} \phi_\nu(\boldsymbol{X}[j]|\boldsymbol{X}[\nu, j-1]), \tag{2.21}$$

where $\phi_\nu(\boldsymbol{X}[j]|\boldsymbol{X}[\nu, j-1])$ denotes the post-change conditional distribution of $\boldsymbol{X}[j]$ given past observations (see (2.10)) and changepoint equal to ν, and where $\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i, j-1]) \triangleq 1$ for $i \geq j$. Then, the GLR test statistic between the two hypotheses can be written as

$$W'_G[k] \triangleq \max_{1\leq i\leq k} \sum_{j=i}^{k} \log \frac{\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i, j-1])}{g(\boldsymbol{X}[j])} = \max_{1\leq i\leq k} \log \mathcal{L}(k, i), \tag{2.22}$$

where

$$\mathcal{L}(k, \nu) \triangleq \frac{h_\nu(\boldsymbol{X}[1, k])}{g(\boldsymbol{X}[1, k])} \tag{2.23}$$

denotes the likelihood ratio of $\boldsymbol{X}[1, k]$ between the hypothesis that the anomaly appears at time ν and the hypothesis that the anomaly never appears, and the corresponding stopping rule for threshold $b > 0$ is given by

$$\tau'_G(b) \triangleq \inf\{k \geq 1 : W'_G[k] \geq b\}. \tag{2.24}$$

Although the conditional pdf $\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1])$ in (2.22) can be calculated recursively (as shown below), to compute $W'_G[k]$, the number of quantities that need to be stored scales with time k, which is not feasible for a real-time algorithm. Thus, to design an implementable GLR test we consider a *windowed* version of $W'_G[k]$. Denote the windowed version of the GLR statistic in (2.22) by

$$W_G[k] \triangleq \max_{k-\eta \leq i \leq k} \sum_{j=i}^{k} \log \frac{\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1])}{g(\boldsymbol{X}[j])}, \tag{2.25}$$

where η denotes window length. In addition, define the corresponding stopping time by

$$\tau_G(b) \triangleq \inf\{k \geq 1 : W_G[k] \geq b\}. \tag{2.26}$$

As will be observed later, the window length η needs to scale with threshold b (and as a result γ), and also depends on the sensor data-generating distributions.

Note that for fixed j, i such that $j > i$, $\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1])$ can be calculated recursively by a standard Bayesian update. In particular, by using the Bayes rule it can be easily shown that for $\boldsymbol{E} \in \mathcal{E}$

$$\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1]) = \sum_{\boldsymbol{E}' \in \mathcal{E}} p_{\boldsymbol{E}'}(\boldsymbol{X}[j]) \mathbb{P}_i(\boldsymbol{S}[j] = \boldsymbol{E}'|\boldsymbol{X}[i,j-1]), \tag{2.27}$$

$$\mathbb{P}_i(\boldsymbol{S}[j] = \boldsymbol{E}|\boldsymbol{X}[i,j-1]) = \sum_{\boldsymbol{E}' \in \mathcal{E}} \mathbb{P}_i(\boldsymbol{S}[j-1] = \boldsymbol{E}'|\boldsymbol{X}[i,j-1]) \lambda_{\boldsymbol{E}',\boldsymbol{E}}, \tag{2.28}$$

$$\begin{aligned} &\mathbb{P}_i(\boldsymbol{S}[j-1] = \boldsymbol{E}|\boldsymbol{X}[i,j-1]) = \mathbb{P}_i(\boldsymbol{S}[j-1] = \boldsymbol{E}|\boldsymbol{X}[i,j-2], \boldsymbol{X}[j-1]) \\ &= \frac{\mathbb{P}_i(\boldsymbol{S}[j-1] = \boldsymbol{E}|\boldsymbol{X}[i,j-2]) p_{\boldsymbol{E}}(\boldsymbol{X}[j-1])}{\sum_{\boldsymbol{E}' \in \mathcal{E}} \mathbb{P}_i(\boldsymbol{S}[j-1] = \boldsymbol{E}'|\boldsymbol{X}[i,j-2]) p_{\boldsymbol{E}'}(\boldsymbol{X}[j-1])}, \end{aligned} \tag{2.29}$$

where the recursion is initialized with the stationary probability of the

DTMC:

$$\mathbb{P}_i(\boldsymbol{S}[i] = \boldsymbol{E}) \triangleq \alpha_{\boldsymbol{E}}, \tag{2.30}$$

for all $\boldsymbol{E} \in \mathcal{E}$.

For the Windowed-GLR test, we can establish a lower bound on the MTFA.

Lemma 1. *For the stopping rule defined in* (2.25) *and* (2.26)*, the MTFA can be lower bounded as follows:*

$$\mathbb{E}_\infty[\tau_G(b)] \geq e^b. \tag{2.31}$$

Proof. The details of the proof can be found in Appendix A.1. □

2.5 Asymptotic Optimality of the Windowed-GLR Test

In this section, we present the first-order asymptotic optimality of the Windowed-GLR detection procedure. Before proceeding to the main results, we define the following effective KL number:

$$J \triangleq \lim_{k\to\infty} \frac{1}{k} \frac{h_1(\boldsymbol{X}[1,k])}{g(\boldsymbol{X}[1,k])} = \mathbb{E}_1\left[\log \frac{h_1(\boldsymbol{X}[1], \boldsymbol{X}[2])}{g(\boldsymbol{X}[1,2])}\right], \tag{2.32}$$

where the underlying probability measure is $\mathbb{P}_1(\cdot)$. Such a limit is assumed to exist almost surely with $0 < J < \infty$, which is the case if the non-anomalous and anomalous data-generating distributions are distinct at each node. KL-type quantities often play a crucial role in characterizing the asymptotic performance of QCD procedures, as will be seen in the remainder of this dissertation.

We first present the universal lower bound on the CADD (and thus on the WADD) for any stopping rule τ that satisfies the false alarm constraint.

Theorem 1. *Consider the QCD problem outlined in Sections 2.1-2.3. If conditions C.1 and C.2 are satisfied, then as $\gamma \to \infty$,*

$$\inf_{\tau \in \mathcal{C}_\gamma} \text{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \text{CADD}(\tau) \geq \frac{\log \gamma}{J}(1 + o(1)). \tag{2.33}$$

Proof. The details of the proof can be found in Appendix A.2. □

Next, we establish an asymptotic upper bound on the WADD and CADD of the Windowed-GLR test introduced in (2.25) and (2.26).

Theorem 2. *Consider the stopping rule defined in* (2.25) *and* (2.26)*. Consider the window length* $\eta \triangleq \eta(b)$ *such that*

$$\liminf_{b\to\infty} \frac{\eta(b)}{b} > \frac{1}{J}. \tag{2.34}$$

Then, under conditions C.1 and C.2, we have that as $b \to \infty$*,*

$$\mathrm{CADD}(\tau_G(b)) \leq \mathrm{WADD}(\tau_G(b)) \leq \frac{b}{J}(1 + o(1)). \tag{2.35}$$

Proof. The details of the proof can be found in Appendix A.3. □

Finally, the following theorem demonstrates the asymptotic optimality of the Windowed-GLR test, which follows directly from Lemma 1 and Theorems 1 and 2.

Theorem 3. *Consider the stopping rule defined in* (2.25) *and* (2.26) *with* $b = \log\gamma$ *and* η *chosen to satisfy*

$$\liminf_{\gamma\to\infty} \frac{\eta(\gamma)}{\log\gamma} > \frac{1}{J}. \tag{2.36}$$

Then, under conditions C.1 and C.2, the windowed-GLR test is first-order asymptotically optimal under both (2.16) *and* (2.17)*, i.e., as* $\gamma \to \infty$

$$\mathrm{WADD}(\tau_G(\log\gamma)) \sim \mathrm{CADD}(\tau_G(\log\gamma)) \sim \frac{\log\gamma}{J}. \tag{2.37}$$

Proof. The result follows directly from Lemma 1 and Theorems 1 and 2. □

2.6 Alternative Detection Schemes

In this section, we develop several alternative algorithms for the problem of Markov anomaly detection, and derive lower bounds on their MTFAs. We first design a *Dynamic Shiryaev-Roberts* (D-S-R) algorithm by modeling the changepoint as a geometric random variable with parameter ρ, and then letting $\rho \to 0$. The advantage of the D-S-R algorithm is that it can be updated recursively. We then develop a QCD algorithm with recursive changepoint estimation. This test recursively estimates the unknown changepoint, and then constructs a CUSUM-type algorithm using the estimated changepoint. Finally, we design a *Mixture-CUSUM* algorithm, which is applicable for the case where the Markov transition probabilities are unknown.

2.6.1 Dynamic Shiryaev-Roberts Test

For our first alternative test, we initially assume that the changepoint is a geometric random variable with parameter ρ. We denote this geometric changepoint by Γ. Specifically,

$$\mathbb{P}(\Gamma = i) = \rho(1-\rho)^{i-1}, \; i \geq 1. \tag{2.38}$$

In the following, we will show how we design a recursive test under such a Bayesian framework. We will further let $\rho \to 0$ so that the designed algorithm does not depend on ρ, and can be applied to the minimax setting studied in this chapter, where the changepoint is deterministic and unknown.

Under the Bayesian assumption of the changepoint, we introduce one addition state $\boldsymbol{E}_0 \notin \mathcal{E}$ to denote the state where there is no anomaly in the network. Then, the transition from the pre-change mode to the post-change mode can be represented by the transition from state $\boldsymbol{E}_0$ to any state $\boldsymbol{E} \in \mathcal{E}$. Specifically, for $\boldsymbol{E} \in \mathcal{E}$ we denote by $\lambda_{\boldsymbol{E}_0,\boldsymbol{E}}$ the probability that the anomaly first emerges at initial node placement given by $\boldsymbol{E}$, i.e.,

$$\mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{S}[k-1] = \boldsymbol{E}_0) \triangleq \lambda_{\boldsymbol{E}_0,\boldsymbol{E}}. \tag{2.39}$$

It is clear that

$$\rho = \sum_{\boldsymbol{E} \in \mathcal{E}} \lambda_{\boldsymbol{E}_0,\boldsymbol{E}}. \tag{2.40}$$

We further note that $\lambda_{\boldsymbol{E},\boldsymbol{E}_0} \triangleq 0$, for any $\boldsymbol{E} \in \mathcal{E}$, and $\lambda_{\boldsymbol{E}_0,\boldsymbol{E}_0} = 1 - \rho$. For any

$\boldsymbol{E} \in \mathcal{E} \cup \{\boldsymbol{E}_0\}$, and $k \geq 1$, define by

$$q_{\boldsymbol{E}}[k] \triangleq \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k]) \tag{2.41}$$

the posterior probability that the network is at state $\boldsymbol{E}$ at time k. A natural way to construct a test is to compare with a threshold the posterior probability that the network is in the pre-change state.

Note that $q_{\boldsymbol{E}}[k]$ can be updated recursively. In particular, for any $\boldsymbol{E} \in \mathcal{E} \cup \{\boldsymbol{E}_0\}$, by the Bayes rule we have that

$$\begin{aligned}
q_{\boldsymbol{E}}[k] &= \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k-1], \boldsymbol{X}[k]) \\
&= \frac{\mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k-1], \boldsymbol{X}[k]) p(\boldsymbol{X}[k] | \boldsymbol{X}[1,k-1])}{p(\boldsymbol{X}[k] | \boldsymbol{X}[1,k-1])} \\
&= \frac{\mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k-1]) p(\boldsymbol{X}[k] | \boldsymbol{S}[k] = \boldsymbol{E}, \boldsymbol{X}[1,k-1])}{\sum_{\boldsymbol{E}' \in \mathcal{E}} p(\boldsymbol{X}[k], \boldsymbol{S}[k] = \boldsymbol{E}' | \boldsymbol{X}[1,k-1])} \\
&= \frac{\mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k-1]) p(\boldsymbol{X}[k] | \boldsymbol{S}[k] = \boldsymbol{E}, \boldsymbol{X}[1,k-1])}{\sum_{\boldsymbol{E}' \in \mathcal{E}} \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E}' | \boldsymbol{X}[1,k-1]) p(\boldsymbol{X}[k] | \boldsymbol{S}[k] = \boldsymbol{E}', \boldsymbol{X}[1,k-1])} \\
&= \frac{B_{\boldsymbol{E}}[k]}{\sum_{\boldsymbol{E}' \in \mathcal{E}} B_{\boldsymbol{E}'}[k]},
\end{aligned} \tag{2.42}$$

where $p(\cdot|\cdot)$ denotes the conditional probability density function of $\boldsymbol{X}[k]$ and for $\boldsymbol{E} \in \mathcal{E}$

$$\begin{aligned}
B_{\boldsymbol{E}}[k] &\triangleq \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k-1]) p(\boldsymbol{X}[k] | \boldsymbol{S}[k] = \boldsymbol{E}, \boldsymbol{X}[1,k-1]) \\
&= \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E} | \boldsymbol{X}[1,k-1]) p_{\boldsymbol{E}}(\boldsymbol{X}[k]).
\end{aligned} \tag{2.43}$$

We then compute $B_{\boldsymbol{E}}[k]$ as follows:

$$\begin{aligned}
B_{\boldsymbol{E}}[k] &= \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E}|\boldsymbol{X}[1,k-1])p_{\boldsymbol{E}}(\boldsymbol{X}[k]) \\
&= \sum_{\boldsymbol{E}' \in \mathcal{E} \cup \{\boldsymbol{E}_0\}} \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E}, \boldsymbol{S}[k-1] = \boldsymbol{E}'|\boldsymbol{X}[1,k-1])p_{\boldsymbol{E}}(\boldsymbol{X}[k]) \\
&= \sum_{\boldsymbol{E}' \in \mathcal{E} \cup \{\boldsymbol{E}_0\}} \mathbb{P}(\boldsymbol{S}[k-1] = \boldsymbol{E}'|\boldsymbol{X}[1,k-1]) \\
&\qquad\qquad \cdot \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E}|\boldsymbol{S}[k-1] = \boldsymbol{E}', \boldsymbol{X}[1,k-1])p_{\boldsymbol{E}}(\boldsymbol{X}[k]) \\
&= \sum_{\boldsymbol{E}' \in \mathcal{E} \cup \{\boldsymbol{E}_0\}} \mathbb{P}(\boldsymbol{S}[k-1] = \boldsymbol{E}'|\boldsymbol{X}[1,k-1]) \\
&\qquad\qquad \cdot \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E}|\boldsymbol{S}[k-1] = \boldsymbol{E}')p_{\boldsymbol{E}}(\boldsymbol{X}[k]) \\
&= \left[\sum_{\boldsymbol{E}' \in \mathcal{E} \cup \{\boldsymbol{E}_0\}} q_{\boldsymbol{E}'}[k-1]\lambda_{\boldsymbol{E}',\boldsymbol{E}}\right] p_{\boldsymbol{E}}(\boldsymbol{X}[k]).
\end{aligned} \tag{2.44}$$

Combining (2.42) and (2.44) implies that $q_{\boldsymbol{E}}[k]$ can be updated recursively.

We note that for $\boldsymbol{E}' \in \mathcal{E}$

$$\begin{aligned}
B_{\boldsymbol{E}}[k] &= \left[\sum_{\boldsymbol{E}' \in \mathcal{E} \cup \{\boldsymbol{E}_0\}} q_{\boldsymbol{E}'}[k-1]\lambda_{\boldsymbol{E}',\boldsymbol{E}}\right] p_{\boldsymbol{E}}(\boldsymbol{X}[k]) \\
&= \left[q_{\boldsymbol{E}_0}[k-1]\lambda_{\boldsymbol{E}_0,\boldsymbol{E}} + \sum_{\boldsymbol{E}' \in \mathcal{E}} q_{\boldsymbol{E}'}[k-1]\lambda_{\boldsymbol{E}',\boldsymbol{E}}\right] p_{\boldsymbol{E}}(\boldsymbol{X}[k]).
\end{aligned} \tag{2.45}$$

Furthermore, for $\boldsymbol{E} = \boldsymbol{E}_0$ we have that

$$\begin{aligned}
B_{\boldsymbol{E}_0}[k] &= \mathbb{P}(\boldsymbol{S}[k] = \boldsymbol{E}_0|\boldsymbol{X}[1,k-1])p_{\boldsymbol{E}_0}(\boldsymbol{X}[k]) \\
&= q_{\boldsymbol{E}_0}[k-1]\lambda_{\boldsymbol{E}_0,\boldsymbol{E}_0} g(\boldsymbol{X}[k]).
\end{aligned} \tag{2.46}$$

The recursion is initialized with $q_{\boldsymbol{E}_0}[0] \triangleq 1$ and $q_{\boldsymbol{E}}[0] \triangleq 0$ for all $\boldsymbol{E} \in \mathcal{E}$.

We further define the following invertible mapping:

$$w_{\boldsymbol{E}}[k] = \frac{q_{\boldsymbol{E}}[k]}{\rho q_{\boldsymbol{E}_0}[k]} \Leftrightarrow q_{\boldsymbol{E}}[k] = \frac{w_{\boldsymbol{E}}[k]}{\sum_{\boldsymbol{E}' \in \mathcal{E} \cup \{\boldsymbol{E}_0\}} w_{\boldsymbol{E}'}[k]}. \tag{2.47}$$

It then follows that

$$q_{\boldsymbol{E}_0}[k] = \frac{1}{1 + \left[\rho \sum_{\boldsymbol{E}' \in \mathcal{E}} w_{\boldsymbol{E}'[k]}\right]}, \tag{2.48}$$

where $w_{\boldsymbol{E}}[k]$ can be computed recursively by

$$w_{\boldsymbol{E}}[k] = \frac{\left[\frac{\lambda_{\boldsymbol{E}_0,\boldsymbol{E}}}{\rho} + \sum_{\boldsymbol{E}' \in \mathcal{E}} w_{\boldsymbol{E}'}[k-1]\lambda_{\boldsymbol{E}',\boldsymbol{E}}\right] p_{\boldsymbol{E}}(\boldsymbol{X}[k])}{\lambda_{\boldsymbol{E}_0,\boldsymbol{E}_0} g(\boldsymbol{X}[k])}, \tag{2.49}$$

with the following priors: $w_{\boldsymbol{E}_0}[k] \triangleq 1/\rho$ and $w_{\boldsymbol{E}}[0] \triangleq 0$, $\boldsymbol{E} \in \mathcal{E}$. From (2.48), it follows that comparing $q_{\boldsymbol{E}_0}[k]$ to a threshold b' is equivalent to comparing

$$\sum_{\boldsymbol{E}' \in \mathcal{E}} w_{\boldsymbol{E}'[k]} \tag{2.50}$$

to a threshold $(1/b' - 1)/\rho$.

To obtain a test that does not depend on ρ and can be applied to the non-Bayesian setting, we take the limit $\rho \to 0$. In particular, we assume that as $\rho \to 0$,

$$\frac{\lambda_{\boldsymbol{E}_0,\boldsymbol{E}}}{\rho} \to \alpha_{\boldsymbol{E}} \tag{2.51}$$

for all $\boldsymbol{E} \in \mathcal{E}$. Practically, this means that the changepoint is treated as an unknown but deterministic variable, and that after the change occurs, the initial location of the anomaly is distributed according to $\boldsymbol{\alpha}$. As a result, if we define

$$r_{\boldsymbol{E}}[k] = \lim_{\rho \to 0} w_{\boldsymbol{E}}[k],$$

for $\boldsymbol{E} \in \mathcal{E}$, then the recursion of $r_{\boldsymbol{E}}[k]$ is

$$r_{\boldsymbol{E}}[k] = \left[\alpha_{\boldsymbol{E}} + \sum_{\boldsymbol{E}' \in \mathcal{E}} r_{\boldsymbol{E}'}[k-1]\lambda_{\boldsymbol{E}',\boldsymbol{E}}\right] \frac{p_{\boldsymbol{E}}(\boldsymbol{X}[k])}{g(\boldsymbol{X}[k])} \tag{2.52}$$

with $r_{\boldsymbol{E}}[0] \triangleq 0$. Define the test statistic

$$W_{SR}[k] \triangleq \sum_{\boldsymbol{E} \in \mathcal{E}} r_{\boldsymbol{E}}[k]. \tag{2.53}$$

The corresponding stopping rule is then given by

$$\tau_{SR}(b) \triangleq \inf \{k \geq 1 : \log W_{SR}[k] \geq b\} . \tag{2.54}$$

This detection scheme involves calculating a test statistic for each possible set of anomalous nodes. At each time k, the test statistic for one possible anomaly allocation is calculated by first weighing the test statistics of all the possible sets of anomalous nodes at the previous time instant according to the corresponding transition probabilities, and then multiplying the likelihood ratio of the sample taken by the anomalous nodes at that specific allocation. Hence, the knowledge of the transition probabilities is needed in order to implement this test.

We note that the D-S-R algorithm is developed by letting $\rho \to 0$. Such a changepoint can be intuitively interpreted as a "uniformly" distributed random variable on the entire time scale. Therefore, this algorithm may not perform as well as the Windowed-GLR test under both Lorden's and Pollak's criteria, since both criteria are defined for the worst-case scenario over all possible changepoints.

Next, we derive a lower bound on the MTFA for the D-S-R algorithm.

Lemma 2. *For the stopping rule defined in* (2.52) - (2.54)*, the MTFA can be lower bounded as follows:*

$$\mathbb{E}_\infty[\tau_{SR}(b)] \geq e^b. \tag{2.55}$$

Proof. The details of the proof can be found in Appendix A.4. □

2.6.2 QCD Algorithm with Recursive Changepoint Estimation

In the Windowed-GLR test, the changepoint is implicitly estimated by the *maximum-likelihood* approach over a finite window. The estimation does not have a recursive form, and hence is not as computationally efficient, which is why a windowed approach is used. An interesting question is whether we can design a test that can recursively and inherently estimate the changepoint, and then construct a CUSUM-type algorithm using the estimated changepoint.

QCD algorithms based on recursive changepoint estimation were proposed in [47] to solve the semi-parametric QCD problem, and in [48] to solve the composite QCD problem (for prior work in composite QCD see [17] and [22]). The main idea is motivated by the CUSUM algorithm, for which, before the changepoint the test statistic takes values around zero, and therefore an estimate of the changepoint is the last time that the test statistic was equal to zero. Following a similar idea, we design a QCD algorithm with recursive changepoint estimation. In particular, define the following test statistic:

$$W_{CE}[k] \triangleq \max_{\hat{\nu}[k-1] \leq i \leq k+1} \sum_{j=i}^{k} \log \frac{\phi_{\hat{\nu}[k-1]}(\boldsymbol{X}[j]|\boldsymbol{X}[\hat{\nu}[k-1], j-1])}{g(\boldsymbol{X}[j])}, \tag{2.56}$$

where $\hat{\nu}[k]$ denotes the estimate of the changepoint at time k. The estimate of the changepoint is defined by

$$\hat{\nu}[k] = \arg\max_{\hat{\nu}[k-1] \leq i \leq k+1} \sum_{j=i}^{k} \log \frac{\phi_{\hat{\nu}[k-1]}(\boldsymbol{X}[j]|\boldsymbol{X}[\hat{\nu}[k-1], j-1])}{g(\boldsymbol{X}[j])}. \tag{2.57}$$

Following steps similar to those in [47], it can be shown that the detection statistics in (2.56) and (2.57) can be updated recursively as follows:

$$W_{CE}[k] = \left(W_{CE}[k-1] + \log \frac{\phi_{\hat{\nu}[k-1]}(\boldsymbol{X}[k]|\boldsymbol{X}[\hat{\nu}[k-1], k-1])}{g(\boldsymbol{X}[k])} \right)^{+}, \tag{2.58}$$

and

$$\hat{\nu}[k] = \begin{cases} \hat{\nu}[k-1], & W_{CE}[k-1] > 0 \text{ or } \hat{\nu}[k-1] = k, \\ k+1, & \text{else}, \end{cases} \tag{2.59}$$

where $W_{CE}[0] \triangleq 0$ and $\hat{\nu}[0] \triangleq 1$. The corresponding stopping rule is

$$\tau_{CE} = \inf \{ k \geq 1 : W_{CE}[k] \geq b \}. \tag{2.60}$$

The advantage of such a test is that it is an approximation to the GLR test which can be implemented recursively. We now present a lower bound for the MTFA for the algorithm defined in (2.56) - (2.60).

Lemma 3. *For the stopping rule defined in* (2.56) - (2.60)*, the MTFA can* *be lower bounded as follows:*

$$\mathbb{E}_{\infty}[\tau_{CE}(b)] \geq e^{b}. \tag{2.61}$$

Proof. The details of the proof can be found in Appendix A.5. □

Due to the use of the recursive changepoint estimate $\hat{\nu}[k]$, the analysis of the detection delay for this algorithm is challenging. We leave this as an open problem for future research.

2.6.3 Mixture-CUSUM Test

In practice, it might be hard to acquire complete knowledge of the transition probabilities of the DTMC in (2.4). However, it might be possible to have a good estimate of the stationary distribution of the DTMC, e.g., based on symmetries in the network, we may be able to approximate the stationary distribution by a uniform distribution. In this case, we approximate the post-change joint data generating distribution by a mixture of $p_{\boldsymbol{E}}(\boldsymbol{x})$, $\boldsymbol{E} \in \mathcal{E}$, where the weights are the stationary distribution $\boldsymbol{\alpha}$, and construct a CUSUM algorithm that tests the change from the pre-change distribution to the mixture distribution.

In particular, the *Mixture-CUSUM* (M-CUSUM) test statistic for the Markov anomaly case is defined as follows:

$$\begin{aligned} W_{\boldsymbol{\alpha}}[k] &\triangleq \max_{1\leq i\leq k} \sum_{j=i}^{k} \log\left(\sum_{\boldsymbol{E}\in\mathcal{E}} \alpha_{\boldsymbol{E}} \frac{p_{\boldsymbol{E}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}\right) \\ &= \max_{1\leq i\leq k} \sum_{j=i}^{k} \log\left(\sum_{\boldsymbol{E}\in\mathcal{E}} \alpha_{\boldsymbol{E}} \prod_{\ell\in\boldsymbol{E}} \frac{f_{\ell}(X_{\ell}[j])}{g_{\ell}(X_{\ell}[j])}\right). \end{aligned} \tag{2.62}$$

Note that this statistic can be equivalently updated recursively:

$$W_{\boldsymbol{\alpha}}[k] = (W_{\boldsymbol{\alpha}}[k-1])^{+} + \log\left(\sum_{\boldsymbol{E}\in\mathcal{E}} \alpha_{\boldsymbol{E}} \prod_{\ell\in\boldsymbol{E}} \frac{f_{\ell}(X_{\ell}[k])}{g_{\ell}(X_{\ell}[k])}\right), \tag{2.63}$$

with $W_{\boldsymbol{\alpha}}[0] \triangleq 0$. The Mixture-CUSUM stopping rule is

$$\tau_{M}(\boldsymbol{\alpha}, b) \triangleq \inf\left\{k \geq 1 : W_{\boldsymbol{\alpha}}[k] \geq b\right\}. \tag{2.64}$$

Since this test is essentially a CUSUM algorithm that tests a change from the pre-change distribution to a mixture post-change distribution, its MTFA can be lower bounded similarly to the CUSUM algorithm.

Lemma 4. *For the M-CUSUM algorithm defined in* (2.62) - (2.64)*, the MTFA can be lower bounded as follows for any* $\boldsymbol{\alpha} \in \mathcal{A}$*:*

$$\mathbb{E}_\infty[\tau_M(\boldsymbol{\alpha}, b)] \geq e^b. \tag{2.65}$$

Proof. The result follows directly from the lower bound on the MTFA for the CUSUM algorithm (see [17], [18] and [22]). □

Since the M-CUSUM algorithm only employs the stationary distribution of the DTMC, we might expect a loss in performance compared to the other algorithms that make use of the entire transition matrix. However, as will be seen in Section 2.8, the M-CUSUM test performs competitively with the presented asymptotically optimal algorithms.

2.7 Fuh's Recursive Approximation Algorithm

In this section, we review Fuh's (see [36]) *recursive approximation algorithm*, and instantiate it for our dynamic anomaly detection problem.

As discussed in Section 2.4, the GLR-based test does not admit a recursion. To address this problem, Fuh in [36] approximates the conditional pdf $\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i, j-1])$ in (2.22) using $\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1, j-1])$. Such an approximation inherently uses the likelihood when the changepoint is at time 1 to approximate the likelihood when the changepoint is at ν. In this way, the log-likelihood ratio does not depend on the changepoint ν, and thus the test statistic can be updated recursively. Specifically, the detection statistic of Fuh's recursive approximation test is given by

$$W_F[k] \triangleq \max_{1 \leq i \leq k} \sum_{j=i}^{k} \log \frac{\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1, j-1])}{g(\boldsymbol{X}[j])}. \tag{2.66}$$

Then, $W_F[k]$ can be written recursively as follows:

$$W_F[k] = (W_F[k-1])^+ + \log \frac{\phi_1(\boldsymbol{X}[k]|\boldsymbol{X}[1, k-1])}{g(\boldsymbol{X}[k])}, \tag{2.67}$$

where $W_F[0] \triangleq 0$. The corresponding stopping rule is defined as

$$\tau_F(b) \triangleq \inf\{k \geq 1 : W_F[k] \geq b\}. \tag{2.68}$$

In [36], Fuh used the stationarity properties of Markov chains to prove the first-order asymptotic optimality of τ_F. For completeness, we include his result in the next theorem.

Theorem 4. *([36]) Consider the stopping rule defined in* (2.66) -(2.68) *with* $b = \log \gamma$. *Then we have that*

$$\mathbb{E}_\infty[\tau_F(\log \gamma)] \geq \gamma \tag{2.69}$$

and that as $\gamma \to \infty$

$$\text{WADD}(\tau_F(\log \gamma)) \sim \text{CADD}(\tau_F(\log \gamma)) \sim \frac{\log \gamma}{J}. \tag{2.70}$$

2.8 Numerical Results

In this section, we conduct a numerical study for the Markov dynamic anomaly detection problem. We set $g_\ell = \mathcal{N}(0, 1)$ and $f_\ell = \mathcal{N}(2, 1)$ for all $\ell \in [L]$. We consider different values of network size L, and compare all the algorithms discussed in this section.

For the Windowed-GLR test, the QCD algorithm with recursive change-point estimation and Fuh's recursive approximation test, the worst case detection delay is not necessarily attained at $\nu = 1$ for the WADD or CADD (also see in [38]). As a result, it is difficult to analytically or numerically calculate the worst-case detection delay for these algorithms. For the D-S-R and M-CUSUM tests, the WADD and CADD are attained at $\nu = 1$. For the purpose of illustration, we simulate the average detection delay $\mathbb{E}_\nu [\tau - \nu | \tau \geq \nu]$ for different values of the changepoint ν, which serves as an approximation for the WADD and CADD.

In Fig. 2.1, we evaluate the value of J as a function of the network size L. The KL number J was calculated by the Monte Carlo method according to (2.32). Note that J decreases with network size. This implies that for a large network, the Windowed-GLR test requires a large window size. In Fig. 2.2,

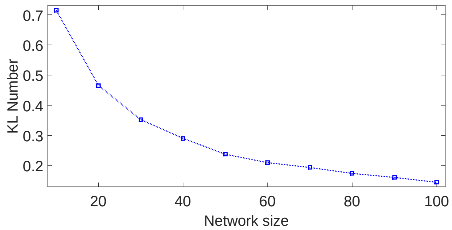


Figure 2.1: J versus L.

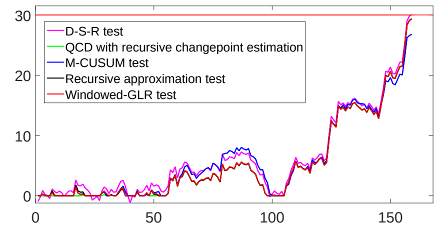


Figure 2.2: Evolution of test statistics.

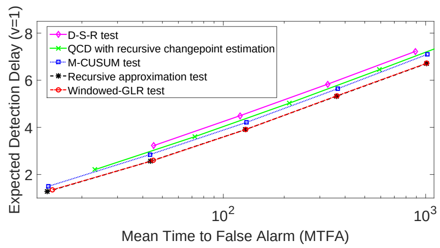


Figure 2.3: $\mathbb{E}_1[\tau - 1|\tau \geq 1]$ versus MTFA for $L = 10$.

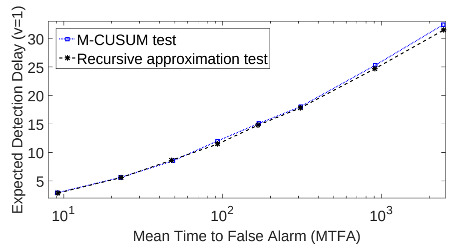


Figure 2.4: $\mathbb{E}_1[\tau - 1|\tau \geq 1]$ versus MTFA for $L = 100$.

we plot the evolution of test statistics for $L = 100$ and $\nu = 120$. It can be seen that the statistics for all the algorithms grow after the changepoint. In Fig. 2.3, we plot the average detection delay vs. MTFA for the algorithms discussed in this chapter for $\nu = 1$, $L = 10$ and $\eta = 30$. Among all the tests, the Windowed-GLR test, Fuh's recursive approximation algorithm, and the M-CUSUM test perform the best. In the remainder of this section, we mainly compare these three algorithms. In Fig. 2.4, we first compare

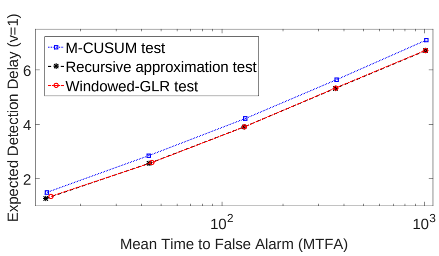


Figure 2.5: $\mathbb{E}_1[\tau - 1|\tau \geq 1]$ versus MTFA for $L = 10$.

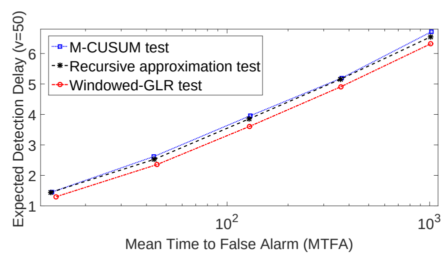


Figure 2.6: $\mathbb{E}_{50}[\tau - 50|\tau \geq 50]$ versus MTFA for $L = 10$.

Fuh's test with the M-CUSUM test for $L = 100$ and $\nu = 1$. We note that although the M-CUSUM algorithm only employs the stationary distribution of the DTMC, and does not use the transition probabilities, it provides very good performance compared to Fuh's recursive approximation test,

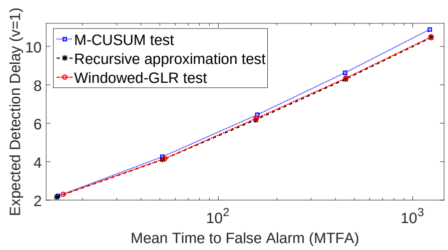


Figure 2.7: $\mathbb{E}_1[\tau - 1|\tau \geq 1]$ versus MTFA for $L = 20$.

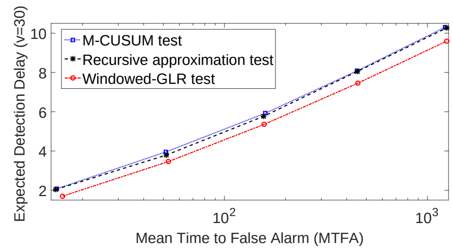


Figure 2.8: $\mathbb{E}_{30}[\tau - 30|\tau \geq 30]$ versus MTFA for $L = 20$.

which is provably first-order asymptotically optimal. Furthermore, Fuh's test can be computationally expensive for a large L, since it requires $O(L^2)$ computations per time step, while the computational complexity for the M-CUSUM algorithm is only $O(L)$. Thus, for large networks, the M-CUSUM test might be a better choice if computational resources are limited. In Fig. 2.5, we repeat the comparison for $L = 10$, $\eta = 30$ and $\nu = 1$ by adding the Windowed-GLR test, and similar observations are obtained. Note that in this case Fuh's recursive test offers identical performance to the Windowed-GLR, since the former inherently assumes that the change occurs at $\nu = 1$. In Fig. 2.6, we further compare Fuh's test and the M-CUSUM test with the Windowed-GLR test for $L = 10$, $\eta = 30$ and $\nu = 50$. Note that although for the case of $\nu = 1$ the Windowed-GLR test has a similar performance to Fuh's algorithm, the Windowed-GLR test performs better for $\nu \neq 1$. This phenomenon is expected since Fuh's test is using the likelihood when $\nu = 1$ as an approximation. Finally, in Figs. 2.7 and 2.8 we compare the three tests for the case of $L = 20$ with $\eta = 50$, $\nu = 1$ and $\nu = 30$, and obtain similar conclusions.

CHAPTER 3

CONSTANT-SIZE WORST-PATH DYNAMIC ANOMALY DETECTION

In this chapter, we study the problem of dynamic anomaly detection in sensor networks under a worst-path setting, for anomalies of constant size. In Chapter 2, we considered the dynamic anomaly detection setting when the anomaly moves according to a DTMC. However, as was mentioned there, assuming knowledge of the transition probabilities of a DTMC, especially for the case of large networks, is a very hard assumption to guarantee in practice. As a result, we need to consider settings where such model knowledge is not needed. Furthermore, we saw in Chapter 2 that the M-CUSUM procedure performs competitively compared to other algorithms requiring complete knowledge of the underlying DTMC. Hence, one interesting question is whether there exists some QCD formulation of the dynamic anomaly setting where the M-CUSUM test is also theoretically justified. To this end, in this chapter we lift the assumption of an underlying DTMC governing the evolution of the anomaly and assume that the path of the anomaly is unknown but deterministic. To balance this lack of knowledge, we introduce a novel modification of Lorden's [17] delay metric used in Chapter 2, that evaluates candidate stopping rules according to their performance on the worst path of the anomaly. Next, we establish that the M-CUSUM test with uniformly chosen weights is exactly optimal for the defined worst-path delay vs. MTFA QCD framework. Furthermore, we show that we can choose the parameters of the M-CUSUM test such that we get a first-order asymptotically optimal procedure even when the sensors are heterogeneous. We conclude the chapter by comparing the M-CUSUM test with other heuristic and oracle tests (oracle algorithms use complete knowledge of the anomaly path) for the case of homogeneous sensors, as well as by investigating the performance loss that we incur when the algorithm parameters are not chosen optimally in the heterogeneous sensors case. This chapter has appeared in part as [49–51].

3.1 Observation Model

In this chapter, we focus on the problem of *constant-size* anomalies; hence, the observation model is the same as that in Section 2.1. In particular, we have that *conditioned* on ν and $\boldsymbol{S}$ the complete statistical model is the following:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]), & 1 \leq k < \nu, \\ p_{\boldsymbol{S}[k]}(\boldsymbol{X}[k]), & k \geq \nu. \end{cases} \tag{3.1}$$

Here we also assume that observations are independent across sensors, and independent across time before the changepoint. Similarly, datapoints sampled before and after the changepoint are assumed to be independent. The main difference here is that the trajectory process $\boldsymbol{S}$ is *unknown* and *deterministic*, and not assumed to evolve according to a DTMC as in Chapter 2. This also implies that observations are independent across time and sensors after the changepoint, *conditioned* on $\boldsymbol{S}$.

Note that the dynamic anomaly QCD problem described in (3.1) can also be posed as the following dynamic composite hypothesis testing problem: at each time instant k, decide between the hypotheses

$$\begin{aligned} \mathscr{H}_0^k &: \text{the anomaly appears at time } \nu > k, \\ \mathscr{H}_{1,\boldsymbol{S}}^k &: \text{the anomaly appears at time } \nu \leq k \text{ and evolves according to } \boldsymbol{S}. \end{aligned} \tag{3.2}$$

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

$$\Gamma_{\boldsymbol{S}}(k,\nu) \triangleq \prod_{j=\nu}^{k} \left(\prod_{\ell \in \boldsymbol{S}[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])} \right) = \prod_{j=\nu}^{k} \Gamma_{\boldsymbol{S}}(j,j). \tag{3.3}$$

3.2 Problem Formulation

Since in this chapter the anomaly trajectory process $\boldsymbol{S}$ is assumed to be deterministic, we modify Lorden's delay metric used in Chapter 2 to evaluate candidate detection schemes according to the anomaly path that maximizes their 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 $\boldsymbol{X}$ consider the following modification of Lorden's WADD metric:

$$\mathrm{WADD}(\tau) \triangleq \sup_{\boldsymbol{S}} \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu^{\boldsymbol{S}} \left[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\right], \tag{3.4}$$

where the convention that $\mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \triangleq 1$ when $\mathbb{P}_\nu^{\boldsymbol{S}}(\tau \geq \nu) = 0$ is used. Note that an additional sup is used to account for the trajectory of the anomaly that maximizes the detection delay of τ. Our goal then is to design a stopping time τ to solve the following constrained optimization problem:

$$\begin{aligned} \min_{\tau} \quad & \mathrm{WADD}(\tau) \\ \text{s.t.} \quad & \tau \in \mathcal{C}_\gamma. \end{aligned} \tag{3.5}$$

3.3 Randomized Anomaly Allocation Model

Before proceeding to the presentation of our main theoretical results, it is necessary to introduce another statistical model that plays an important role in the analysis, as well as in the interpretation of the results in this chapter. In particular, consider an alternate setting to that of (3.1), where at each time instant after the changepoint, the m anomalous nodes are chosen *randomly*. To this end, denote by $\boldsymbol{\alpha} \triangleq \{\alpha_{\boldsymbol{E}} : \boldsymbol{E} \in \mathcal{E}\} \in \mathcal{A}$ 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 sets of anomalous nodes are picked i.i.d. across time. 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 joint pdf after the changepoint is a mixture of pdfs given by

$$\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[k]) \triangleq \sum_{\boldsymbol{E} \in \mathcal{E}} \alpha_{\boldsymbol{E}} p_{\boldsymbol{E}}(\boldsymbol{X}[k]). \tag{3.6}$$

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

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]), & 1 \leq k < \nu, \\ \overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[k]), & k \geq \nu. \end{cases} \tag{3.7}$$

Remark 3. *It is important to note that the model described in* (3.7) *is an intermediate model that is going to be used to facilitate algorithm design and analysis in order to solve* (3.5)*. The main observation model of interest in this chapter is the one outlined in Section* (3.1)*, which is the model that governs data generation in this chapter.*

Similarly to (3.1), we can pose the following dynamic composite hypothesis testing problem corresponding to (3.7): at each time k choose between the hypotheses

$$\begin{aligned} \bar{\mathscr{H}}_0^k &: \text{the anomaly appears at time } \nu > k, \\ \bar{\mathscr{H}}_{1,\boldsymbol{\alpha}}^k &: \text{the anomaly appears at time } \nu \leq k \text{ and is placed according to } \boldsymbol{\alpha}. \end{aligned} \tag{3.8}$$

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

$$\mathcal{L}_{\boldsymbol{\alpha}}(k,\nu) \triangleq \prod_{j=\nu}^{k} \frac{\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])} = \prod_{j=\nu}^{k} \left(\sum_{\boldsymbol{E} \in \mathcal{E}} \alpha_{\boldsymbol{E}} \prod_{\ell \in \boldsymbol{E}} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])} \right) = \prod_{j=\nu}^{k} \mathcal{L}(j,j). \tag{3.9}$$

We also denote the KL divergence between the post- and pre-change distributions in (3.7) by

$$I_{\boldsymbol{\alpha}} \triangleq D(\overline{p}_{\boldsymbol{\alpha}} \| g) = \overline{\mathbb{E}}_1^{\boldsymbol{\alpha}} \left[\log \frac{\overline{p}_{\boldsymbol{\alpha}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right], \tag{3.10}$$

where $\overline{\mathbb{E}}_\nu^{\boldsymbol{\alpha}}[\cdot]$ denotes the expectation when the underlying statistical model is that of (3.7) with changepoint being equal to ν and the anomaly placed randomly according to $\boldsymbol{\alpha}$.

Note that the model in (3.7) characterizes a different QCD problem compared to the one described in (3.1) - (3.5), one in which the pre- and post-change pdfs are completely specified. 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 (3.7) by

$$\overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau) \triangleq \sup_{\nu \geq 1} \operatorname{ess\,sup} \overline{\mathbb{E}}_{\nu}^{\boldsymbol{\alpha}}[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]]. \tag{3.11}$$

Here, we also use the convention that $\overline{\mathbb{E}}_{\nu}^{\boldsymbol{\alpha}}[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]] \triangleq 1$ when $\overline{\mathbb{P}}_{\nu}^{\boldsymbol{\alpha}}(\tau \geq \nu) = 0$. Since both the pre- and post-change joint pdfs for the QCD problem presented in (3.7) - (3.11) are completely specified, the classical CUSUM test studied in [17–19, 22] can be directly applied to solve this QCD problem exactly [19]. In the remainder of this chapter, we show that solving the QCD problem in (3.7) - (3.11) for a carefully chosen $\boldsymbol{\alpha}$, which depends on the data generating distributions of the sensors, leads to a solution to the QCD problem of interest described in eqs. (3.1) - (3.5).

3.4 Mixture-CUSUM Test

In Chapter 2, we numerically established that the M-CUSUM test with mixture weights chosen according to the stationary probabilities performs competitively compared to provably asymptotically optimal procedures that require complete model knowledge and are more computationally demanding (see Section 2.8). In this chapter, we study the application of the M-CUSUM test in the worst-path dynamic anomaly detection setting. To this end, for $\boldsymbol{\lambda} \in \mathcal{A}$, consider the following M-CUSUM test statistic:

$$W_{\boldsymbol{\lambda}}[k] \triangleq \max_{1 \leq i \leq k} \mathcal{L}_{\boldsymbol{\lambda}}(k, i) \tag{3.12}$$

with the corresponding stopping time

$$\tau_M(\boldsymbol{\lambda}, b) \triangleq \inf\left\{k \geq 1 : W_{\boldsymbol{\lambda}}[k] \geq e^b\right\}. \tag{3.13}$$

It can be easily established (see, e.g., [11]) that for any $\boldsymbol{\lambda} \in \mathcal{A}$ the test statistic in (3.12) can be computed recursively as

$$W_{\boldsymbol{\lambda}}[k] = \max\{W_{\boldsymbol{\lambda}}[k-1], 1\}\mathcal{L}_{\boldsymbol{\lambda}}(k, k), \tag{3.14}$$

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

Remark 4. *Note that the version of the M-CUSUM test described in eqs.* (3.12) - (3.14) *is equivalent to the one in* (2.62) - (2.64)*, and is only used in this chapter to facilitate a clean and concise analysis of the performance of the algorithm.*

From the exact optimality of the CUSUM test [19] it follows that the M-CUSUM test presented in eqs. (3.12) - (3.14) is the exact solution to the QCD problem detailed in (3.7) - (3.11) for $\gamma > 0$ when $\boldsymbol{\alpha} = \boldsymbol{\lambda}$, if b is chosen such that $\mathbb{E}_\infty[\tau_M(\boldsymbol{\alpha}, b)] = \gamma$. In the remainder of this chapter, we establish that by choosing $\boldsymbol{\lambda}$ accordingly the M-CUSUM procedure is also an exact solution to (3.5) 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 Sections 3.1 - 3.2 and 3.3, and exploiting tools used for the analysis of the CUSUM test in [19, 22]. Before proceeding to establish the optimality properties of the M-CUSUM test, we present an important theorem relating the detection delay metrics (3.4) and (3.11), introduced in Sections 3.2 and 3.3 respectively.

Theorem 5. *Let $\gamma > 0$ and $\boldsymbol{\alpha} \in \mathcal{A}$. Consider the QCD problems outlined in Sections 3.1 -3.2 and 3.3. Consider the stopping rule defined in* (3.12) - (3.14) *with b chosen such that $\mathbb{E}_\infty[\tau_M(\boldsymbol{\alpha}, b)] = \gamma$. We have that*

$$\mathrm{WADD}(\tau_M(\boldsymbol{\alpha}, b)) \geq \inf_{\tau \in C_\gamma} \mathrm{WADD}(\tau) \geq \overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau_M(\boldsymbol{\alpha}, b)). \tag{3.15}$$

Proof. The details of the proof can be found in Appendix B.2. □

3.5 Homogeneous Sensor Network Case

We begin by considering 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 in this case is symmetric, an intuitive weight choice for the M-CUSUM test of (3.12)-(3.14) is one where all the weights are equal. This then implies that by the symmetry of the statistical model, as well as the resulting symmetry of the detection procedure with respect to the placement of the anomaly, placing the anomaly randomly or according to the worst-path approach will not lead to a different detection delay. In particular, we have the following lemma:

Lemma 5. *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* $\boldsymbol{\lambda}_U \triangleq \left[\binom{L}{m}, \ldots, \binom{L}{m}\right]^\top$ *be the uniform M-CUSUM weights vector. For any threshold* $b > 0$ *and any* $\boldsymbol{\alpha} \in \mathcal{A}$ *we have that*

$$\mathrm{WADD}(\tau_M(\boldsymbol{\lambda}_U, b)) = \overline{\mathrm{WADD}}_{\boldsymbol{\alpha}}(\tau_M(\boldsymbol{\lambda}_U, b)). \tag{3.16}$$

Proof. The details of the proof can be found in Appendix B.3. □

By using Theorem 5 and Lemma 5 we can establish the exact optimality of the M-CUSUM test with uniform weights for the case of a homogeneous sensor network.

Theorem 6. *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* $\gamma > 0$. *The M-CUSUM test with uniform weights* $\boldsymbol{\lambda} = \boldsymbol{\lambda}_U \triangleq \left[\binom{L}{m}, \ldots, \binom{L}{m}\right]^\top$ *and threshold* b *chosen such that* $\mathbb{E}_\infty[\tau_M(\boldsymbol{\lambda}_U, b)] = \gamma$ *is exactly optimal with respect to* (3.5), *i.e.,*

$$\mathrm{WADD}(\tau_M(\boldsymbol{\lambda}_U, b)) = \inf_{\tau \in C_\gamma} \mathrm{WADD}(\tau). \tag{3.17}$$

Proof. The result follows directly by combining Theorem 5 and Lemma 5. □

Theorem 6 implies that, for the case of homogeneous sensors, the M-CUSUM test that solves the QCD problem of eqs. (3.7) - (3.11) for a uniform pmf $\boldsymbol{\alpha} = \boldsymbol{\lambda}_U$ is also the exact solution to (3.1) - (3.5). Next, we investigate whether a similar result holds for the general case of heterogeneous networks.

3.6 Heterogeneous Sensor Network Case

In Section 3.5, we saw how the symmetry of a homogeneous sensor network can facilitate the construction of an exactly optimal test with respect to (3.5). However, in the case of a heterogeneous sensor network, such a symmetry no longer holds, and a result similar to Lemma 5 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 by exploiting an asymptotic type of symmetry that is related to the expected drift of the test statistic.

3.6.1 Universal Asymptotic Lower Bound on the WADD

We begin our analysis for the heterogeneous sensor network setting by presenting an asymptotic lower bound on WADD for stopping times in $\mathcal{C}_\gamma$. Our lower bound is derived by using Theorem 5 together with the asymptotic lower bound on $\overline{\text{WADD}}$ [17, 22]. In particular, note that the inequalities in Theorem 5 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 $\overline{\text{WADD}}$. To this end, define the minimizer of the effective KL divergence $I_{\boldsymbol{\alpha}}$ by

$$\overset{*}{\boldsymbol{\alpha}} \triangleq \arg\min_{\boldsymbol{\alpha} \in \mathcal{A}} I_{\boldsymbol{\alpha}}. \tag{3.18}$$

It can be shown that $I_{\boldsymbol{\alpha}}$ is strictly convex with respect to $\boldsymbol{\alpha}$, hence, such a minimizer is uniquely defined. As a result, we can define the minimum value of $I_{\boldsymbol{\alpha}}$ by

$$\overset{*}{I} \triangleq I_{\overset{*}{\boldsymbol{\alpha}}}. \tag{3.19}$$

We then have the following theorem:

Theorem 7. *Let $\overset{*}{I}$ be defined as in* (3.19). *Consider the QCD problem outlined in Sections 3.1 - 3.2. We then have that as $\gamma \to \infty$*

$$\inf_{\tau \in \mathcal{C}_\gamma} \text{WADD}(\tau) \geq \frac{\log \gamma}{\overset{*}{I}}(1 + o(1)). \tag{3.20}$$

Proof. By Theorem 5 we have that for any $\boldsymbol{\alpha} \in \mathcal{A}$ and any $\gamma > 0$

$$\inf_{\tau \in \mathcal{C}_\gamma} \text{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \overline{\text{WADD}}_{\boldsymbol{\alpha}}(\tau), \tag{3.21}$$

which implies that the inequality also holds for $\boldsymbol{\alpha} = \overset{*}{\boldsymbol{\alpha}}$, i.e., as $\gamma \to \infty$

$$\inf_{\tau \in \mathcal{C}_\gamma} \text{WADD}(\tau) \geq \inf_{\tau \in \mathcal{C}_\gamma} \overline{\text{WADD}}_{\overset{*}{\boldsymbol{\alpha}}}(\tau) \sim \frac{\log \gamma}{\overset{*}{I}}, \tag{3.22}$$

where the asymptotic delay approximation follows from the asymptotic analysis of the CUSUM test [17, 22]. □

3.6.2 Asymptotic Upper Bound on the WADD of the M-CUSUM Test

Although deriving a lower bound on WADD is similar for both homogeneous and heterogeneous sensor networks (Theorem 5), upper bounding WADD in the latter case for arbitrary $\boldsymbol{\lambda}$ is nontrivial. To find the weight choice of the M-CUSUM test that results in an asymptotically optimal test, it is important to further investigate the minimization of $I_{\boldsymbol{\alpha}}$. To this end, we present the following lemma:

Lemma 6. *Let* $\overset{*}{\boldsymbol{\alpha}}$ *be defined as in* (3.18)*. We then have the following:*

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

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

$$\mathbb{E}_{p_{\boldsymbol{E}}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] = \mathbb{E}_{p_{\boldsymbol{E}'}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] \tag{3.23}$$

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 $2 \leq \|\overset{*}{\boldsymbol{\alpha}}\|_0 < |\mathcal{E}|$ *(boundary-point minimum), let* $\mathcal{E}' \triangleq \{\boldsymbol{E} \in \mathcal{E} : \overset{*}{\alpha}_{\boldsymbol{E}} > 0\}$ *the subset of vectors in* $\mathcal{E}$ *for which non-zero weights are assigned in* $\overset{*}{\boldsymbol{\alpha}}$*. We then have that for all* $\boldsymbol{E}$*,* $\boldsymbol{E}' \in \mathcal{E}'$ *eq.* (3.23) *holds. Furthermore, we have that for all* $\boldsymbol{B} \in \mathcal{E}'$*,* $\boldsymbol{B}' \in \mathcal{E} \setminus \mathcal{E}'$

$$\mathbb{E}_{p_{\boldsymbol{B}'}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] > \mathbb{E}_{p_{\boldsymbol{B}}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right]. \tag{3.24}$$

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

Proof. The details of the proof can be found in Appendix B.4. □

By exploiting the properties presented in Lemma 6, we derive an asymptotic upper bound on $\mathrm{WADD}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b))$. In particular, we have the following theorem:

Theorem 8. *Let $\overset{*}{\boldsymbol{\alpha}}$ be defined as in* (3.18). *Assume that*

$$\max_{\boldsymbol{E} \in \mathcal{E}} \mathbb{E}_{p_{\boldsymbol{E}}} \left[\left(\log \frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right)^2 \right] < \infty. \tag{3.25}$$

Consider the stopping rule defined in (3.12) - (3.14). *We then have that as $b \to \infty$*

$$\mathrm{WADD}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)) \leq \frac{b}{\overset{*}{I}}(1 + o(1)). \tag{3.26}$$

Proof. The details of the proof can be found in Appendix B.5. □

3.6.3 Asymptotic Optimality of the M-CUSUM Test

By combining Theorems 7 with 8 we can establish the asymptotic optimality of the M-CUSUM test for weight choice $\boldsymbol{\lambda} = \overset{*}{\boldsymbol{\alpha}}$.

Theorem 9. *Let $\overset{*}{\boldsymbol{\alpha}}$, $\overset{*}{I}$ be defined as in* (3.18) *and* (3.19) *respectively, and assume that*

$$\max_{\boldsymbol{E} \in \mathcal{E}} \mathbb{E}_{p_{\boldsymbol{E}}} \left[\left(\log \frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right)^2 \right] < \infty. \tag{3.27}$$

Consider the stopping rule defined in (3.12) - (3.14). *We then have that:*

i) For any $\gamma > 0$, $\boldsymbol{\lambda}$

$$\mathbb{E}_\infty[\tau_M(\boldsymbol{\lambda}, \log \gamma)] \geq \gamma. \tag{3.28}$$

ii) The M-CUSUM test with $\boldsymbol{\lambda} = \overset{*}{\boldsymbol{\alpha}}$ is first-order asymptotically optimal under (3.5), *i.e., as $\gamma \to \infty$*

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}(\tau) \sim \mathrm{WADD}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, \log \gamma)) \sim \frac{\log \gamma}{\overset{*}{I}}. \tag{3.29}$$

Proof. i) Follows directly from the MTFA analysis of the CUSUM test [17,22].

ii) Follows from i) and Theorems 7 and 8. □

Essentially, Theorem 9 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 (3.7) - (3.11) for said $\boldsymbol{\alpha}$ exactly is also asymptotically optimal with respect to (3.1) - (3.5). This $\boldsymbol{\alpha}$ is the unique minimizer of the KL divergence in (3.10).

The asymptotic optimality of the M-CUSUM test with weights given by $\overset{*}{\boldsymbol{\alpha}}$ can be intuitively explained through Lemma 6. In particular, since a larger γ implies a larger threshold, if we consider the logarithm of the M-CUSUM test statistic in (3.12) (equivalent M-CUSUM form studied in (2.62)), 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 6 we know that choosing $\boldsymbol{\lambda} = \overset{*}{\boldsymbol{\alpha}}$ implies that the drift of the statistic is equal among a specific subset of anomaly placements. In other words, the M-CUSUM test with $\boldsymbol{\lambda} = \overset{*}{\boldsymbol{\alpha}}$ is an equalizer rule with respect to the drift of the test statistic among different anomaly allocations for a subset of $\mathcal{E}$. Furthermore, as we see in Lemma 6, all other placements of anomalous nodes lead to a larger drift and hence do not play a role asymptotically due to the worst-path aspect of the delay. This equalization of slopes is the reason that the M-CUSUM test with optimal weights matches the universally best delay asymptotically.

Remark 5. *It should be noted that the first-order asymptotic optimality results in this chapter also hold if we use a worst-path version of Pollak's detection delay [18]. In particular, for stopping time* τ *define the detection delay*

$$\mathrm{CADD}(\tau) \triangleq \sup_{\boldsymbol{S}} \sup_{\nu \geq 1} \mathbb{E}_{\nu}^{\boldsymbol{S}} \left[\tau - \nu | \tau \geq \nu \right]. \tag{3.30}$$

By deriving an inequality connecting CADD *and the corresponding Pollak's delay metric under the randomized anomaly allocation model of Sec. 3.3., 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. As a result, Theorem 9 also holds when* WADD *is replaced by* CADD. *However, it is not clear whether the M-CUSUM test is exactly optimal with* *respect to Pollak's criterion for the case of homogeneous sensor networks, since the exact optimality of the CUSUM test for Pollak's formulation in the classical single-sensor QCD setting has not been established. Hence, it can not be exploited to prove the exact optimality of the M-CUSUM test, as was done in Appendix B.2.*

3.7 Numerical Results

In this section, we present numerical results for the worst-path dynamic anomaly QCD problem studied in this chapter 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.

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

$$\mathbb{E}_{\infty}\left[\sum_{\ell=1}^{L} \log \frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)D(f\|g)\right] = -mD(f\|g) < 0$$
$$\mathbb{E}_1^{\boldsymbol{S}}\left[\sum_{\ell=1}^{L} \log \frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)D(f\|g)\right] = mD(f\|g) > 0.$$

This suggests that the following Naive-CUSUM (N-CUSUM) test may be a candidate test for detecting the distribution change described in (3.1). In particular, consider the test described by the following recursion:

$$W_N[k] \triangleq (W_N[k-1])^+ + \sum_{\ell=1}^{L} \log \frac{f(X_\ell[k])}{g(X_\ell[k])} + (L-m)D(f\|g) \tag{3.31}$$

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

$$\tau_N \triangleq \inf\left\{k \geq 1 : W_N[k] \geq b\right\}.$$

Although the N-CUSUM test can be employed to detect the anomaly reasonably well due to the statistic $W_N[k]$ having the right drift behavior before and after the change, it does not necessarily solve the QCD problem in (3.5).

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 occurred, 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:

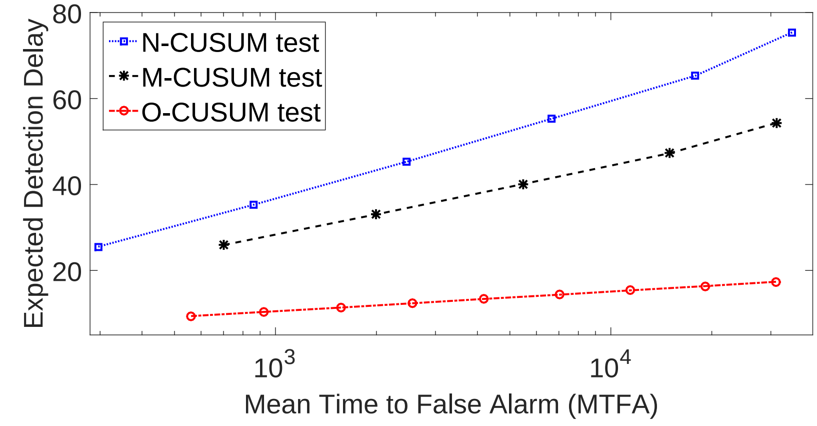
$$W_O[k] = (W_O[k-1])^+ + \log\left(\prod_{\ell \in \boldsymbol{S}[k]} \frac{f(X_\ell[k])}{g(X_\ell[k])}\right) \tag{3.32}$$

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

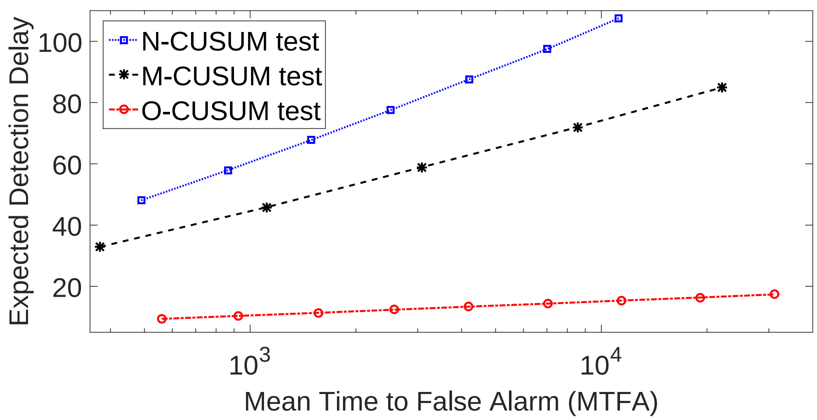
$$\tau_O \triangleq \inf\left\{k \geq 1 : W_O[k] \geq b\right\}. \tag{3.33}$$

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 Figs. 3.1(a), 3.1(b) and 3.2(a) we compare the M-CUSUM test with the N-CUSUM test and the O-CUSUM test for network sizes $L = 5$, $L = 10$ and $L = 20$. 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 Figs. 3.1(a), 3.1(b) and 3.2(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 (3.5). 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 $\boldsymbol{S}$. We also note that as L increases the performance gap between the O-CUSUM test and the M-CUSUM test increases. This is because as the network size increases the "noise" that is introduced in the M-CUSUM test due to nodes that are not anomalous also increases. This is not the case for the O-CUSUM test, since this scheme inherently assumes complete knowledge of the anomalous nodes. In Fig. 3.2(b), we evaluate the performance of our proposed M-CUSUM test for different values of L. We note that as



(a) WADD versus MTFA for $L = 5$, $m = 1$.

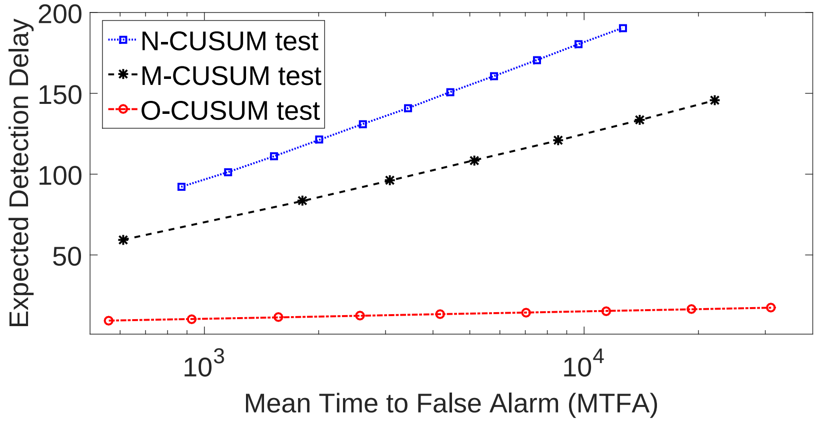


(b) WADD versus MTFA for $L = 10$, $m = 1$.

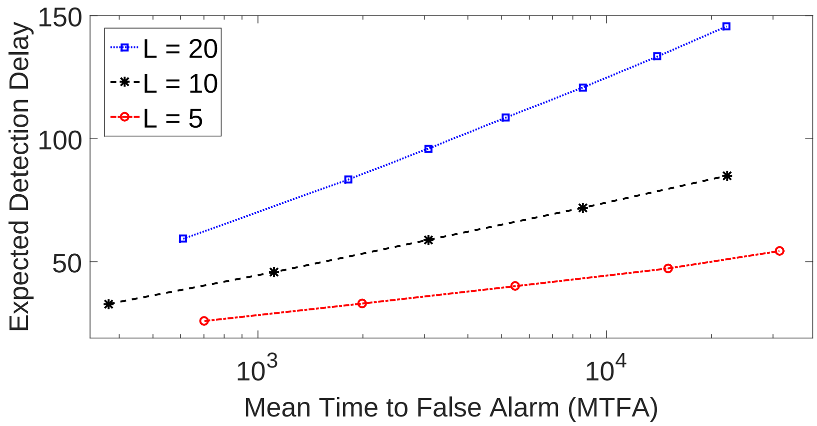
Figure 3.1: WADD versus MTFA for homogeneous sensor networks.

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.

For the case of a heterogeneous sensor network, we compare three versions of the test introduced in eqs. (3.12) - (3.14): the first version ("Optimal weights" in Fig. 3.3) uses the optimal weights $\overset{*}{\boldsymbol{\alpha}}$ to achieve a uniform average statistic drift among anomaly placements (see Lemma 6); the second and third versions ("Non-optimal weights 1" and "Non-optimal weights 2" in Figs. 3.3) 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 eq. (B.62). Note that, according to (B.62), each derivative



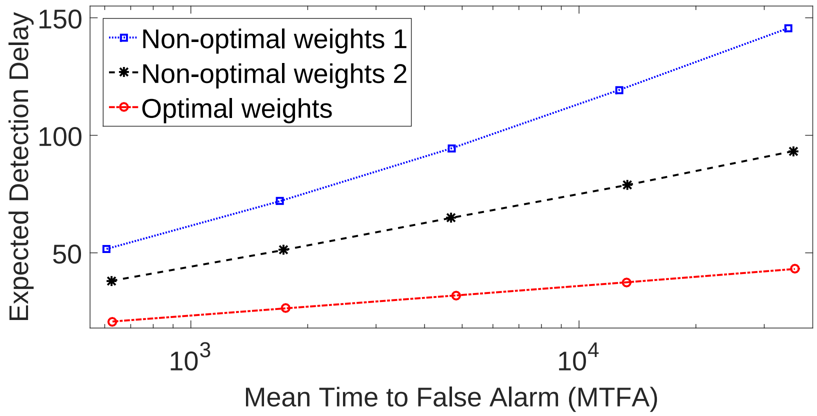
(a) WADD versus MTFA for $L = 20$, $m = 1$.



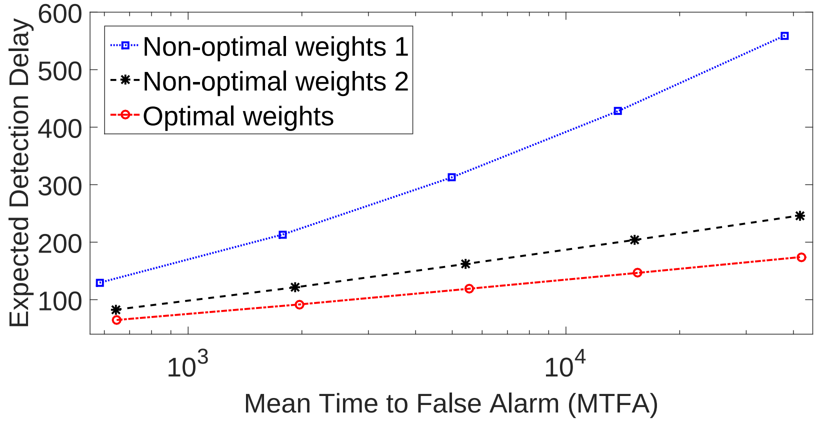
(b) WADD versus MTFA for the M-CUSUM when $m = 1$ and for different L values.

Figure 3.2: WADD versus MTFA for homogeneous sensor networks.

is equal to a difference of two expected values, which we calculate using Monte Carlo methods. Furthermore, 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 the nodes (in this case node since $m = 1$) that correspond to the smallest 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 $f_\ell = \mathcal{N}(\theta_\ell, 1)$ with $\boldsymbol{\theta} = [1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9]^\top$ denoting the vector of the mean



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



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

Figure 3.3: WADD versus MTFA for heterogeneous sensor networks.

values of the anomalous distributions. The results can be seen in Fig. 3.3(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-optimal weights 1" the smallest drift corresponds to placing the anomaly at sensor 2, corresponding to an approximate slope of 0.029, and for the case of "Non-optimal weights 2" at sensor 5, with an approximate slope of 0.065. We see that the Mixture-CUSUM test using the optimal weights $\overset{*}{\boldsymbol{\alpha}}$ 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. 3.3(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-optimal weights 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-optimal weights 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-optimal weights 1" to correspond to the case of uniform weights. As a result, the gap between the blue and red lines in Fig. 3.3(b) captures the loss we suffer if we wrongly make the assumption that the sensors of the network are homogeneous.

CHAPTER 4

VARYING-SIZE WORST-PATH DYNAMIC ANOMALY DETECTION

In Chapter 3, we studied the problem of sequentially detecting dynamic anomalies of constant size under a worst-path delay metric. Although the extension to detecting anomalies of varying size for the Markov setting can be easily achieved by modifying the structure of the DTMC, as mentioned in Chapter 2, generalizing the results of Chapter 3 to the varying-size case is non-trivial. In this chapter, we study the problem of worst-path dynamic anomaly detection for anomalies of varying size. Our main assumption is that the dynamic anomaly to be detected evolves not only in space, but also in size through a series of phases. Every phase corresponds to a specific anomaly size, with the final phase referred to as the *persistent phase* and the intermediate phases as the *transient phases*. We frame this varying-size dynamic anomaly detection problem under a worst-path setting by extending the delay metric introduced in Chapter 3 to account for the presence of transient phases. Similarly to Chapter 3, where the constant-size dynamic anomaly QCD problem was solved by associating it with an instance of the classical QCD setting, in this chapter we use results from transient QCD [35]. In particular, we establish that a version of the *Weighted Dynamic-CUSUM* procedure, that solves a specific instance of the transient QCD problem studied in [35], is asymptotically optimal for the dynamic anomaly detection problem of interest in this chapter. The proposed detection scheme involves a set of parameters to be chosen by minimizing KL numbers, such as in Chapter 3. The main difference here is that there is a different mixture weight vector per post-change phase, resulting from a minimization of a specific KL divergence per phase. Finally, we numerically evaluate our proposed procedure for different cases, such as different network sizes, different degrees of model knowledge, and for the case of optimal vs. non-optimal parameter choice. This chapter has appeared in part as [52, 53].

4.1 Observation Model

We begin this chapter by outlining the observation model for the case of varying-size dynamic anomalies. As in the case of Chapters 2 and 3, before the emergence of the dynamic anomaly in the network, it is assumed that sensors generate data i.i.d. across time with respect to their non-anomalous distributions. As a result, the joint pdf of the observations before the anomaly emerges is given by

$$g(\boldsymbol{X}[k]) \triangleq \prod_{\ell=1}^{L} g_\ell(X_\ell[k]). \tag{4.1}$$

At some *unknown* and *deterministic* changepoint $\nu_1 \geq 1$, a dynamic 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 dynamic anomaly size. Phase $i \in [K]$ is assumed to begin at an *unknown* and *deterministic* changepoint ν_i, where $\nu_i \geq \nu_{i'}$ for $i > i'$. As a result, the duration of the i-th transient phase is given by

$$d_i \triangleq \nu_{i+1} - \nu_i \tag{4.2}$$

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 unknown and deterministic. 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 $i \in [K]$. 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 of $\boldsymbol{S}^{(i)}[k]$ corresponding to all anomaly allocations for an anomaly of size $m^{(i)}$. Note that there are $|\mathcal{E}^{(i)}| = \binom{L}{m^{(i)}}$ such positions (here also, 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 $\{\nu_i\}_{i=1}^k$, 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 $\nu_{K+1} \triangleq \infty$)

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

As a result, *conditioned* on $\{\nu_i\}_{i=1}^K$ and $\boldsymbol{S}$ the observations are independent and the complete statistical model is the following:

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]), & 1 \leq k < \nu_1, \\ p_{\boldsymbol{S}^{(i)}[k]}(\boldsymbol{X}[k]), & \nu_i \leq k < \nu_{i+1}, \end{cases} \tag{4.3}$$

for $i \in [K]$.

4.2 Problem Formulation

Our goal in this chapter is to design a detection algorithm to detect the abrupt distribution change occurring at time ν_1, described in (4.3), as quickly as possible, subject to FA constraints. To this end, we use a generalization of the delay metric introduced in Chapter 3 to account for the presence of transient phases during which the anomaly changes in size. Explicitly, denote by $\mathbb{E}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}[\cdot]$ the expectation under the statistical model in (4.3) for fixed ν_1, $\boldsymbol{d}$ and $\boldsymbol{S}$. Then, for any stopping rule τ adapted to $\boldsymbol{X}$ and for vector $\boldsymbol{d}$ define the following delay metric:

$$\text{WADD}_{\boldsymbol{d}}(\tau) \triangleq \sup_{\boldsymbol{S}} \sup_{\nu_1 \geq 1} \operatorname{ess\,sup} \mathbb{E}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}} \left[\tau - \nu_1 + 1 | \tau \geq \nu_1, \boldsymbol{X}[1, \nu_1 - 1] \right], \tag{4.4}$$

where the convention that $\mathbb{E}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}[\tau - \nu_1 + 1 | \tau \geq \nu_1, \boldsymbol{X}[1, \nu - 1]] \triangleq 1$ when $\mathbb{P}_{\nu_1,\boldsymbol{d}}^{\boldsymbol{S}}(\tau \geq \nu_1) = 0$ is used. Note that the proposed detection delay depends on $\boldsymbol{d}$, since different phase durations imply different probability distributions across time, hence different delay for τ. Our goal in this chapter is to design a stopping procedure τ that solves the following stochastic optimization problem for $\gamma > 0$:

$$\begin{aligned} \min_{\tau} \quad & \mathrm{WADD}_{\boldsymbol{d}}(\tau) \\ \text{s.t.} \quad & \tau \in \mathcal{C}_\gamma \end{aligned} \tag{4.5}$$

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

4.3 Randomized Anomaly Allocation Model

As in Chapter 3, in this section, we introduce an alternative statistical model to that in (4.3), only used as an intermediate tool that will play an important role in the presentation of our results, as well as in the analysis. More explicitly, consider the case of a dynamic anomaly that at each phase i affects one of the sets of sensors in $\mathcal{E}^{(i)}$ at random. To this end, denote by $\boldsymbol{\alpha}^{(i)} \triangleq \left\{\alpha^{(i)}_{\boldsymbol{E}} : \boldsymbol{E} \in \mathcal{E}^{(i)}\right\} \in \mathcal{A}^{(i)}$ the 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 during phase i (here, $\mathcal{A}^{(i)}$ denotes the simplex of all probability vectors of dimension $|\mathcal{E}^{(i)}|$). 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 by $\boldsymbol{\alpha} \triangleq \{\boldsymbol{\alpha}^{(i)}\}_{i=1}^K$ the set of the aforementioned pmfs for all phases. According to this randomized allocation model, we have that the joint pdf before the emergence of the anomaly is going to be the same with the pre-change joint pdf in (4.3). In addition, after the emergence of the anomaly we have that the joint pdf of the observations at phase i is completely specified and given by

$$\overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}}(\boldsymbol{X}[k]) \triangleq \sum_{\boldsymbol{E} \in \mathcal{E}^{(i)}} \alpha^{(i)}_{\boldsymbol{E}} p_{\boldsymbol{E}}(\boldsymbol{X}[k]). \tag{4.6}$$

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

$$\boldsymbol{X}[k] \sim \begin{cases} g(\boldsymbol{X}[k]), & 1 \le k < \nu_1, \\ \overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}}(\boldsymbol{X}[k]), & \nu_i \le k < \nu_{i+1}, \end{cases} \tag{4.7}$$

for $i \in [K]$. Furthermore, for fixed $\boldsymbol{\alpha}$ define the KL divergence between the joint pdf at phase i and the non-anomalous joint pdf $g(\boldsymbol{x})$ by

$$I^{(i)}_{\boldsymbol{\alpha}^{(i)}} \triangleq D(\overline{p}^{(i)}_{\boldsymbol{\alpha}^{(i)}} \| g). \tag{4.8}$$

Note that (4.7) corresponds to a transient QCD problem, as described in [35], since the pre-change and post-change pdfs are completely specified. This transient 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 (4.7) for fixed ν_1, $\boldsymbol{\alpha}$, $\boldsymbol{d}$. Then, for stopping time τ define the detection delay corresponding to the QCD problem detailed in (4.7) by

$$\overline{\text{WADD}}_{\boldsymbol{\alpha},\boldsymbol{d}}(\tau) \triangleq \sup_{\nu_1 \geq 1} \operatorname{ess\,sup} \overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu_1,\boldsymbol{d}}[\tau - \nu_1 + 1 | \tau \geq \nu_1, \boldsymbol{X}[\nu_1 - 1]], \tag{4.9}$$

where the convention that $\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu_1,\boldsymbol{d}}[\tau - \nu_1 + 1 | \tau \geq \nu_1, \boldsymbol{X}[\nu_1 - 1]] \triangleq 1$ 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 (4.7)-(4.9) can be solved by using the *Weighted Dynamic-CUSUM* (WD-CUSUM) test proposed in [35], which is first-order asymptotically optimal. However, it is not clear whether this solution coincides with the solution to (4.5). In the remainder of the chapter, we show that solving the transient QCD problem in (4.7) - (4.9) for a specific choice of pmfs in $\boldsymbol{\alpha}$ will lead to the solution of the initial worst-path problem described in (4.5).

4.4 Mixture-WD-CUSUM Test

In this section, we present the *Mixture-WD-CUSUM* (M-WD-CUSUM) test that solves the transient QCD problem introduced in eq. (4.7) - (4.9). In particular, consider the following M-WD-CUSUM test statistic:

$$\Omega_{\boldsymbol{\lambda}}[k] = \max\{\Omega^{(1)}_{\boldsymbol{\lambda}}[k], \ldots, \Omega^{(K)}_{\boldsymbol{\lambda}}[k], 0\}, \tag{4.10}$$

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

$$\Omega^{(i)}_{\boldsymbol{\lambda}}[k] = \max_{0 \leq j \leq i} \left(\Omega^{(j)}_{\boldsymbol{\lambda}}[k-1] + \sum_{r=j}^{i-1} \log \rho_r \right) + \log \frac{\overline{p}^{(i)}_{\boldsymbol{\lambda}^{(i)}}(\boldsymbol{X}[k])}{g(\boldsymbol{X}[k])} + \log(1 - \rho_i), \tag{4.11}$$

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

$$\tau_{\Omega}(\boldsymbol{\lambda}, b) \triangleq \inf\{k \geq 1 : \Omega_{\boldsymbol{\lambda}}[k] \geq b\}. \tag{4.12}$$

From the results in [35], the M-WD-CUSUM test presented in (4.10) - (4.12) is first-order asymptotically optimal with respect to the transient QCD problem in (4.7) - (4.9) when $\boldsymbol{\alpha} = \boldsymbol{\lambda}$ for carefully chosen ρ_i parameters . Explicitly, the ρ_i parameters in (4.11) 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 asymptotically optimal test to be derived. More details regarding choosing the ρ_i parameters will be given in the subsequent analysis, and can also be found in [35]. In the remainder of the chapter, we leverage the results of [35] to establish that choosing $\boldsymbol{\alpha}$ accordingly will lead to the first-order asymptotic optimality of the M-WD-CUSUM with respect to (4.5).

4.5 Universal Asymptotic Lower Bound on the WADD

We begin our analysis by presenting an asymptotic lower bound on WADD for stopping times in $\mathcal{C}_{\gamma}$. As in Chapter 3, our lower bound is based on a lemma connecting the delays in eqs. (4.4) and (4.9). In particular, our first lemma implies 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 $\boldsymbol{\alpha}$. We use this lemma and the asymptotic results in [35] to derive the tightest asymptotic lower bound on WADD. In particular, the lemma is as follows:

Lemma 7. *Consider the QCD problems outlined in Sections 4.1 - 4.2 and 4.3. For any stopping time τ, vector of pmfs $\boldsymbol{\alpha}$ and $\boldsymbol{d}$ we have that*

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau) \geq \overline{\mathrm{WADD}}_{\boldsymbol{\alpha},\boldsymbol{d}}(\tau). \tag{4.13}$$

Proof. The details of the proof can be found in Appendix C.3. □

Remark 6. *Note that Lemma 7 also holds for the case of constant-size worst-path dynamic anomaly detection studied in Chapter 3. However, in Chapter 3* *we used Lemma 5 mainly because it facilitates the proof of the exact optimality of the M-CUSUM test in the homogeneous sensors case.*

Since the results in [35] provide a universal asymptotic lower bound on $\overline{\text{WADD}}$ for any $\boldsymbol{\alpha}$, an asymptotic lower bound on WADD then follows directly from Lemma 7. However, since the asymptotic rate in the lower bound of $\overline{\text{WADD}}$ is a function of the KL numbers defined in (4.8), we need to choose the pmfs in $\boldsymbol{\alpha}$ to get the tightest lower bound on WADD. To this end, define

$$\overset{*}{\boldsymbol{\alpha}}^{(i)} \triangleq \underset{\boldsymbol{\alpha}^{(i)} \in \mathcal{A}^{(i)}}{\arg\min}\, I^{(i)}_{\boldsymbol{\alpha}^{(i)}}. \tag{4.14}$$

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

$$\overset{*}{I}^{(i)} \triangleq I^{(i)}_{\overset{*}{\boldsymbol{\alpha}}^{(i)}}. \tag{4.15}$$

To ensure that the transient phases play a non-trivial role asymptotically, the durations of the transient phases need to scale to infinity accordingly 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 $\gamma \to \infty$

$$d_i \sim c_i \frac{\log \gamma}{\overset{*}{I}^{(i)}}, \tag{4.16}$$

where $d_K \triangleq \infty$. 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 (also see [35]). The specific choice of KL numbers as a scaling coefficient in (4.16) will imply that the universal lower bound and upper bound on the delay of the proposed test will match, as will be noted in the upper bound analysis. To this end, we have the following theorem for the lower bound:

Theorem 10. *Consider the QCD problem defined in Sections 4.1 and 4.2. Assume that* (4.16) *holds. Furthermore, define* $h \triangleq \min\{j \in [K] : \sum_{i=1}^{j} c_i \geq$

1}. *We then have that as* $\gamma \to \infty$

$$\inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}_{\boldsymbol{d}}(\tau) \geq \log \gamma \left(\sum_{i=1}^{h-1} \frac{c_i}{\overset{*}{I}^{(i)}} + \frac{1 - \sum_{i=1}^{h-1} c_i}{\overset{*}{I}^{(h)}} \right) (1 - o(1)). \tag{4.17}$$

Proof. The result follows directly by applying Lemma 7 for $\boldsymbol{\alpha} = \overset{*}{\boldsymbol{\alpha}}$ and using Theorem 5 of [35] to lower bound $\overline{\mathrm{WADD}}_{\overset{*}{\boldsymbol{\alpha}},\boldsymbol{d}}(\tau)$. □

4.6 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 M-WD-CUSUM algorithm. The asymptotic upper bound is based on exploiting the upper bound analysis in [35] and [50]. For the asymptotic upper bound analysis to be non-trivial we need to assume that the transient durations scale accordingly to threshold b. In particular, assume that there exist constants $c_i' \in [0, \infty) \cup \{\infty\}$, $i \in [K-1]$ such that

$$d_i \sim c_i' \frac{b}{\overset{*}{I}^{(i)}}. \tag{4.18}$$

Furthermore, we need to choose the parameters $\rho_i, i \in [K-1]$ in the M-WD-CUSUM test such that their effect is asymptotically negligible [35]. In particular, assume that ρ_i can be chosen such that as $b \to \infty$

$$\rho_i \to 0, \text{ and } -\frac{\log \rho_i}{b} \to 0, \tag{4.19}$$

for $i \in [K-1]$. We then have the following asymptotic upper bound:

Theorem 11. *Consider the QCD problem defined in Sections 4.1 and 4.2. Suppose b and $\rho_i, i \in [K-1]$ are chosen such that* (4.18) *and* (4.19) *hold. Assume that*

$$\max_{i \in [K]} \max_{\boldsymbol{E} \in \mathcal{E}^{(i)}} \mathbb{E}_{p_{\boldsymbol{E}}} \left[\left(\log \frac{\overline{p}^{(i)}_{\overset{*}{\boldsymbol{\alpha}}^{(i)}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])} \right)^2 \right] < \infty. \tag{4.20}$$

Furthermore, define $h' \triangleq \min\{j \in [K] : \sum_{i=1}^{j} c_i' \geq 1\}$.

Consider the stopping rule defined in (4.10) - (4.12)*. We then have that as* $b \to \infty$

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)) \leq b\left(\sum_{i=1}^{h-1} \frac{c'_i}{\overset{*}{I}^{(i)}} + \frac{1-\sum_{i=1}^{h'-1} c'_i}{\overset{*}{I}^{(h')}}\right)(1+o(1)). \tag{4.21}$$

Proof. The details of the proof can be found in Appendix C.4. □

4.7 Asymptotic Optimality of the M-WD-CUSUM Test

By combining Theorems 10 with 11 we can establish the asymptotic optimality of the M-WD-CUSUM when $\boldsymbol{\lambda} = \overset{*}{\boldsymbol{\alpha}}$. In particular, we have the following theorem:

Theorem 12. *Consider the QCD problem defined in Sections 4.1 and 4.2. Assume that*

$$\max_{i \in [K]} \max_{\boldsymbol{E} \in \mathcal{E}^{(i)}} \mathbb{E}_{p_{\boldsymbol{E}}}\left[\left(\log \frac{\overline{p}^{(i)}_{\overset{*}{\boldsymbol{\alpha}}^{(i)}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)^2\right] < \infty. \tag{4.22}$$

Consider the stopping rule defined in (4.10) - (4.12)*. We then have that:*

i) For any $\gamma > 0$, $\boldsymbol{\lambda}$

$$\mathbb{E}_\infty[\tau_\Omega(\boldsymbol{\lambda}, \log\gamma)] \geq \gamma. \tag{4.23}$$

ii) Assume that (4.16) *is satisfied as* $\gamma \to \infty$ *for some* $c_i \in [0, \infty) \cup \{\infty\}$, $i \in [K-1]$*, and that as* $\gamma \to \infty$

$$\rho_i \to 0, \text{ and } -\frac{\log \rho_i}{\log \gamma} \to 0 \tag{4.24}$$

for all $i \in [K-1]$*. Let* $h \triangleq \min\{j \in [K] : \sum_{i=1}^{j} c_i \geq 1\}$*. We then have that the M-WD-CUSUM test with* $\boldsymbol{\lambda} = \overset{*}{\boldsymbol{\alpha}}$ *is first-order asymptotically optimal* *under* (4.5)*, i.e., as* $\gamma \to \infty$

$$\begin{aligned}\mathrm{WADD}_{\boldsymbol{d}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, \log\gamma)) &\sim \inf_{\tau \in \mathcal{C}_\gamma} \mathrm{WADD}_{\boldsymbol{d}}(\tau) \\ &\sim \log\gamma \left(\sum_{i=1}^{h-1} \frac{c_i}{\overset{*}{I}^{(i)}} + \frac{1 - \sum_{i=1}^{h-1} c_i}{\overset{*}{I}^{(h)}} \right). \qquad (4.25)\end{aligned}$$

Proof. i) Follows directly from the MTFA analysis of the WD-CUSUM test [35].

ii) Follows from i) and Theorems 10 and 11, and since for $b = \log\gamma$ we have that $c_i = c'_i$ for all $i \in [K-1]$. □

Remark 7. *Similarly to Chapter 3, the first-order asymptotic optimality results in this chapter also hold if we use a worst-path version of Pollak's detection delay [18]. In particular, for stopping time* τ *and vector of transient durations* $\boldsymbol{d}$ *define the detection delay*

$$\mathrm{CADD}_{\boldsymbol{d}}(\tau) \triangleq \sup_{\boldsymbol{S}} \sup_{\nu_1 \geq 1} \mathbb{E}^{\boldsymbol{S}}_{\nu_1, \boldsymbol{d}} \left[\tau - \nu_1 | \tau \geq \nu_1 \right]. \qquad (4.26)$$

By deriving a lower bound similar to the one in Lemma 7, 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 12 also holds when* WADD *is replaced by* CADD*.*

4.8 Numerical Results

In this section, we numerically evaluate the performance of the proposed M-WD-CUSUM algorithm of (4.10) - (4.12). We consider the case of both homogeneous and heterogeneous sensors. For the case of homogeneous sensors, it can be shown that the optimal weight choice is given by choosing the weights uniformly at each phase [52] . Note that WADD for the proposed test is attained at $\nu_1 = 1$. Furthermore, for the case of heterogeneous sensors, the worst path cannot be specified analytically, as in the case in Chapter 3. 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. We numerically calculate the average statistic slope through Monte Carlo simulations. In addition, we use $\rho_i = \frac{1}{b}$ to guarantee that the conditions in (4.19) are satisfied.

For the case of homogeneous sensors we focus on the case of $g_\ell = \mathcal{N}(0,1)$ and $f_\ell = \mathcal{N}(1,1)$ for $\ell \in [L]$. In Fig. 4.1, 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 (4.11). Furthermore, we see that as the MTFA increases the slopes of the curves decrease gradually. This means that the Mixture-WD-CUSUM is adaptive to each transient phase (also see [35]) since the expected slope of the test statistic increases as the anomaly size increases.

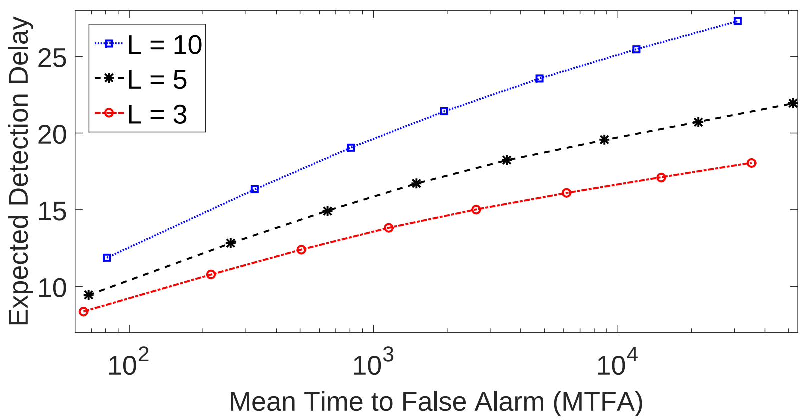


Figure 4.1: WADD versus MTFA for $K = 3$ and varying network sizes.

In Fig. 4.2, 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. 4.3 we evaluate the performance of our proposed detection procedure for the case of a heterogeneous sensor network with $L = K = 5$,

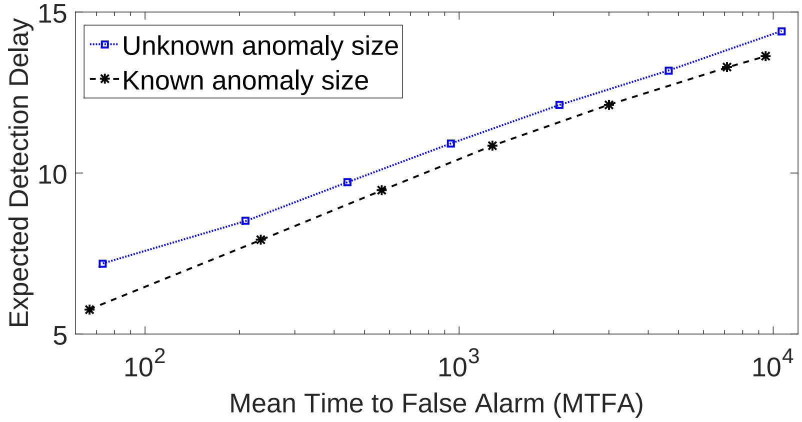


Figure 4.2: WADD versus MTFA comparison between the test that exploits and test that does not exploit complete knowledge of the anomaly size across phases for a homogeneous sensor network.

$m^{(i)} = i$ for $i \in [K]$, $g_\ell = \mathcal{N}(0,1)$ and $f_\ell = \mathcal{N}(\theta_\ell, 1)$ where $\boldsymbol{\theta} = [0.8, 0.8, 1, 1.2, 1.2]^\top$. 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 $f_\ell(\cdot)$ and $g_\ell(\cdot)$ for all $\ell \in [L]$ and chooses the proposed optimal weights 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.

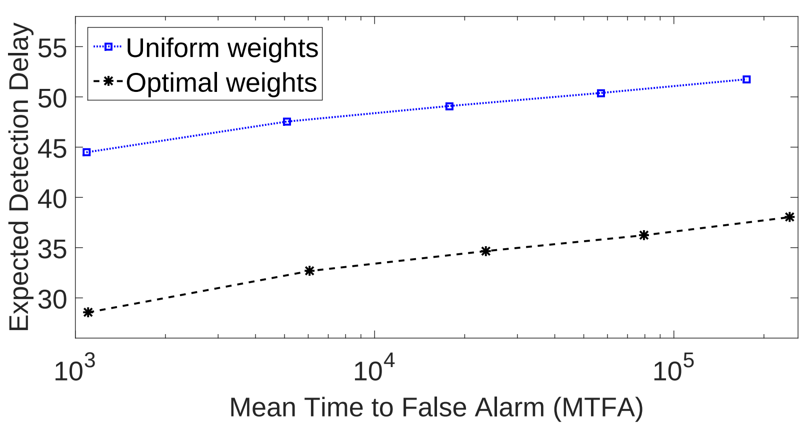


Figure 4.3: WADD versus MTFA comparison between the test that exploits and test that does not exploit complete knowledge of the sensor pdfs for a heterogeneous sensor network.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In this dissertation, we studied the problem of dynamic anomaly detection in sensor networks in a quickest change detection framework. Existing works in the literature of quickest change detection in sensor networks only focus on detecting anomalies that affect sensors persistently. As a result, new formulations are needed to describe problems involving dynamic anomalies, i.e., anomalies affecting the whole network persistently but without necessarily affecting sensors persistently. In this work, we introduced the problem of dynamic anomaly detection in sensor networks and established that algorithms from the literature of classical quickest change detection and transient quickest change detection can be modified to provide tests that offer strong theoretical guarantees, such as exact or asymptotic optimality.

We first studied the setting where an emerging anomaly is modeled as a discrete-time Markov chain in Chapter 2. We constructed the Windowed-GLR test, and established its first-order asymptotic optimality. We also constructed three alternative tests, including the D-S-R test, the QCD test with recursive changepoint estimation and the M-CUSUM test. By conducting comprehensive numerical studies we showed that our Windowed-GLR test provides the best performance in terms of the trade-off between the MTFA and the delay. However, it requires complete knowledge of the probability transitions of the underlying Markov chain and it may suffer from a high computationally complexity especially for large networks. Our proposed M-CUSUM test has a computational complexity of $O(L)$, which is the most efficient among all other tests, and it only requires knowledge of the stationary probabilities of the underlying Markov model, while performing competitively to provably asymptotically optimal tests.

In Chapter 3, we lifted the Markov assumption on the trajectory of the anomaly and focused on a worst-path approach for detecting dynamic anomalies of constant size. To this end, we introduced a modified version of Lorden's detection delay metric [17] that evaluates candidate detection schemes according to their worst performance with respect to the path of the anomaly. We proposed a CUSUM-type test that is an exact solution to the constant-size dynamic anomaly QCD problem for the case of a homogeneous network, and is also first-order asymptotically optimal when applied to a heterogeneous network. We concluded the chapter by conducting a numerical study of the proposed algorithm which included comparisons with heuristic and oracle procedures.

In Chapter 4, we extended the results in Chapter 3 to consider the case of worst-path anomaly detection of anomalies of varying size. As outlined, such an extension is not as straightforward in the worst-path setting as in the Markov setting. For the worst-path varying-size anomaly setting, we made the core assumption that the size of the anomaly evolves in a series of phases, where each phase corresponds to a specific anomaly size. The final phase corresponds to the persistent anomaly size, and the intermediate phases are referred to as transient phases. Under this framework, we established that a WD-CUSUM-type test from the literature of transient QCD [35] is first-order asymptotically optimal. We concluded the chapter by numerically evaluating the proposed detection scheme.

We now discuss possible future directions and interesting problems to be addressed:

1. **Composite/non-parametric setting**: A major assumption throughout this dissertation has been that the decision maker has complete knowledge of the sensor data-generating distributions. This assumption is reasonable for the pre-change distributions of the sensors, since they can be estimated by observing the system operate in the non-anomalous mode. However, in practice the decision maker either has partial knowledge of the anomalous sensor distributions, e.g., up to some unknown parameters (composite setting), or has no knowledge of the anomalous distributions (non-parametric setting). The composite QCD problem has been extensively studied in the literature, albeit in simpler settings than those considered in this dissertation [17, 22]. Although the algorithms proposed in these works are characterized by strong theoretical guarantees, they involve calculating test statistics that cannot be updated recursively. Recently, algorithms that are recursive and

asymptotically optimal have been proposed for the classical composite quickest change detection setting [48]. These tests involve calculating CUSUM-type statistics that use a maximum-likelihood estimator to estimate the parameters of the anomalous distributions. Hence, the QCD problem becomes a problem of joint detection and estimation. To guarantee that the observed test statistic can be updated recursively, these algorithms also include an estimate of the changepoint to be calculated at each time instant. In particular, by exploiting the fact that CUSUM-type statistics take values very close to zero when an anomaly is not present in the network, the maximum-likelihood estimate of the changepoint at each time instant can be shown to be the equal to the last time the test statistic was equal to zero. Similar detection schemes have also been recently used for the problem of semi-parametric QCD [47].

To tackle the problem of lack of post-change model knowledge in our more complicated sensor network setting, our detection procedures have to be enhanced with estimation mechanisms. The main challenges then are the following: (i) similar to [47, 48] the decision maker has no knowledge of whether an anomaly has emerged in the system, hence might use data points from a non-anomalous distribution to estimate the corresponding anomalous distribution; (ii) in the dynamic anomaly setting many of the sensors may generate data points from their non-anomalous distributions even after an anomaly has emerged in the system, hence the decision maker has to construct an estimate of the most probable anomalous nodes at each time instant and only use their measurement for the data-generating distribution estimation. Challenge (i) can be addressed by exploiting recent ideas from [47, 48], since lack of knowledge of the changepoint is apparent in both problems. Regarding Challenge (ii), estimates of the anomalous sensors can be constructed, e.g., in the composite case, by using a generalized likelihood ratio between the anomalous and non-anomalous distribution at each sensor. These generalized likelihood ratios can be ordered and the sensors that correspond to largest likelihood ratio values can then be used as estimates of the most likely anomalous sensors. The mixture weights of the proposed tests can then be calculated by using the estimates of the anomalous distributions. Although these extensions might lead to tractable tests that can be employed in practice, we believe that establishing optimality properties for algorithms of such structure is going to be extremely challenging. However, performance evaluation of the proposed procedures can be achieved by constructing bounds on the asymptotic delay performance, as well as through comprehensive numerical studies with synthetic and real data.

2. **Adversarial setting**: Recently, game-theoretic adversarial settings have been considered in sequential hypothesis testing problems, involving an adversary that can modify the data observed by the decision maker [54–58]. A promising research problem would be to study the extension of algorithms from these works to the QCD setting, and in particular, to the dynamic anomaly detection settings considered in this dissertation.

3. **Distributed dynamic anomaly detection**: In this dissertation, we focus on the problem of dynamic anomaly detection in sensor networks when decision making is done by a centralized decision maker. In practice, anomaly detection using centralized algorithms is difficult to implement due to limited communication bandwidth and long communication distance, especially in large networks. Moreover, in centralized settings data processing is done in one centralized fusion center, often making the complexity of the computations, which usually scales with the network size, unmanageable. To tackle these challenges, distributed algorithms are often employed. In distributed detection settings, part of the information processing is done at the sensor level, by leveraging data obtained from local sensors. The final decision is taken by a fusion center that uses all the results obtained by the data-processing that occurred at each sensor. The problem of detecting a static anomaly that affects all the sensors of the network concurrently by using distributed algorithms has been addressed in [59]. Furthermore, in [33, 34] authors tackled the semi-dynamic anomaly detection problem introduced in [31] by proposing distributed algorithms that require communication between neighboring nodes to be implemented. The complexity of these algorithms at each node was linear to the number of neighbors. An interesting research problem would be to derive distributed versions

of the detection procedures introduced in this dissertation and analyze their performance theoretically and numerically.

4. **Continuous-time setting**: All the problems studied in this dissertation are discrete-time problems, i.e., involve a decision maker that uses data sampled in discrete-time indices. There has been significant work in continuous-time problems in the area of QCD (for a review see, e.g., [12]). It would be interesting to investigate the problem of dynamic anomaly detection in sensor networks in a continuous-time setting.

5. **Data-efficient dynamic anomaly detection**: In many applications anomalies occur rarely, making the cost of sampling observations before an anomaly is present in the system costly. In [60–62], authors studied the problem of data-efficient QCD, where they established that on-off observation control can be introduced in many QCD settings to lead to data-efficient detection schemes. In particular, in [62] the authors considered the problem of sequentially detecting static anomalies in a sensor network by using on-off observation control to limit sampling costs. An interesting research problem would be to incorporate on-off observation control techniques from [60–62] to construct data-efficient algorithms for the problem of dynamic anomaly detection in sensor networks.

APPENDIX A

PROOFS FOR CHAPTER 2

A.1 Proof of Lemma 1

Note that $W'_G[k] \geq W_G[k]$, hence, $\tau_G(b) \geq \tau'_G(b)$ for all b. Therefore,

$$\mathbb{E}_\infty[\tau_G(b)] \geq \mathbb{E}_\infty[\tau'_G(b)]. \tag{A.1}$$

Let

$$W_V[k] \triangleq \sum_{i=1}^{k} \prod_{j=i}^{k} \frac{\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1])}{g(\boldsymbol{X}[j])} = \sum_{i=1}^{k} \mathcal{L}(k,i), \tag{A.2}$$

and

$$\tau_V(b) = \inf\{k \geq 1 : W_V[k] \geq e^b\}. \tag{A.3}$$

Note that

$$\begin{aligned} W_V[k] &= \sum_{i=1}^{k} \frac{\phi_i(\boldsymbol{X}[k]|\boldsymbol{X}[i,k-1])}{g(\boldsymbol{X}[k])} \mathcal{L}(k-1,i) \\ &= \sum_{i=1}^{k-1} \frac{\phi_i(\boldsymbol{X}[k]|\boldsymbol{X}[i,k-1])}{g(\boldsymbol{X}[k])} \mathcal{L}(k-1,i) + \mathcal{L}(k-1,k). \end{aligned} \tag{A.4}$$

Then, from (A.2) and (A.4), we have that

$$\mathbb{E}_\infty[W_V[k]|\boldsymbol{X}[1,k-1]] = 1 + W_V[k-1], \tag{A.5}$$

and

$$\mathbb{E}_\infty[W_V[k]] = k. \tag{A.6}$$

This implies that $\{W_V[k]-k\}_{k=1}^{\infty}$ is a zero-mean martingale under $\mathbb{P}_{\infty}(\cdot)$ [12]. Thus, by the optional sampling theorem (see, e.g., [12]) and the fact that $W_V[k] \geq e^{W'_G[k]}$, we have that

$$\mathbb{E}_{\infty}[\tau_G(b)] \geq \mathbb{E}_{\infty}[\tau'_G(b)] \geq \mathbb{E}_{\infty}[\tau_V(b)] = \mathbb{E}_{\infty}[W_V[\tau_V(b)]] \geq e^b. \tag{A.7}$$

A.2 Proof of Theorem 1

Let $\epsilon > 0$. Define

$$K_\gamma \triangleq \frac{\log \gamma}{J}. \tag{A.8}$$

By Markov's inequality, it follows that

$$\mathbb{E}_\nu[\tau - \nu | \tau \geq \nu] \geq \mathbb{P}_\nu(\tau - \nu \geq K_\gamma(1-\epsilon) | \tau \geq \nu) K_\gamma(1-\epsilon). \tag{A.9}$$

Then, to prove the theorem it suffices to show that for any $\tau \in \mathcal{C}_\gamma$, there exists some $\nu \geq 1$ such that

$$\mathbb{P}_\nu(\nu \leq \tau < \nu + K_\gamma(1-\epsilon) | \tau \geq \nu) = o(1), \tag{A.10}$$

as $\gamma \to \infty$, i.e., that

$$\lim_{\gamma\to\infty} \sup_{\tau \in \mathcal{C}_\gamma} \inf_{\nu \geq 1} \mathbb{P}_\nu(\nu \leq \tau < \nu + K_\gamma(1-\epsilon) | \tau \geq \nu) = 0. \tag{A.11}$$

Define $a \triangleq (1-\epsilon^2)\log\gamma$. Then for any ν, we have that

$$\begin{aligned}
&\mathbb{P}_\nu(\nu \leq \tau < \nu + K_\gamma(1-\epsilon) | \tau \geq \nu) \\
&= \mathbb{P}_\nu\left(\nu \leq \tau < \nu + K_\gamma(1-\epsilon), \mathcal{L}(\tau,\nu) > e^a \middle| \tau \geq \nu\right) \\
&\quad + \mathbb{P}_\nu\left(\nu \leq \tau < \nu + K_\gamma(1-\epsilon), \mathcal{L}(\tau,\nu) \leq e^a \middle| \tau \geq \nu\right).
\end{aligned} \tag{A.12}$$

The first term in (A.12) can be upper bounded as follows:

$$\begin{aligned}
&\mathbb{P}_\nu\left(\nu \le \tau < \nu + K_\gamma(1-\epsilon), \mathcal{L}(\tau,\nu) > e^a \middle| \tau \ge \nu\right)\\
&\overset{(a)}{\le} \mathbb{P}_\nu\left(\max_{\nu \le j < \nu + K_\gamma(1-\epsilon)} \log \mathcal{L}(j,\nu) > a \middle| \tau \ge \nu\right)\\
&\overset{(b)}{=} \mathbb{P}_\nu\left(\max_{\nu \le j < \nu + K_\gamma(1-\epsilon)} \log \mathcal{L}(j,\nu) > a\right)\\
&\overset{(c)}{=} \mathbb{P}_1\left(\frac{\max_{1 \le j < 1 + K_\gamma(1-\epsilon)} \log \mathcal{L}(j,1)}{K_\gamma(1-\epsilon)} > J(1+\epsilon)\right),
\end{aligned} \tag{A.13}$$

where (a) is due to the fact that

$$\{\nu \le \tau < \nu + K_\gamma(1-\epsilon), \log \mathcal{L}(\tau,\nu) > a\} \subseteq \left\{\max_{\nu \le j < \nu + K_\gamma(1-\epsilon)} \log \mathcal{L}(j,\nu) > a\right\}; \tag{A.14}$$

(b) is due to the facts that $\{\tau \ge \nu\} \in \sigma(\boldsymbol{X}[1, \nu-1])$, and the pre- and post-change observations are independent; and (c) follows by the independence between the pre- and post-change observations.

By Lemma A.1 in [63], if

$$\frac{\log \mathcal{L}(k,1)}{k} \xrightarrow[\text{a.s.}]{k\to\infty} J \tag{A.15}$$

under $\mathbb{P}_1(\cdot)$, then it follows that

$$\lim_{\gamma\to\infty} \mathbb{P}_1\left(\frac{\max\limits_{1 \le j < 1 + K_\gamma(1-\epsilon)} \log \mathcal{L}(j,1)}{K_\gamma(1-\epsilon)} > J(1+\epsilon)\right) = 0. \tag{A.16}$$

By Lemma 1. (i) in [39], it follows that if C.1 and C.2 are satisfied, then (A.15) holds, and consequently (A.16) holds.

We then analyze the second term in (A.12). By a change of measure argument similar to the one in [22], we have that there exists $\nu \ge 1$ such that

$$\begin{aligned}
&\mathbb{P}_\nu\left(\nu \le \tau < \nu + K_\gamma(1-\epsilon), \mathcal{L}(\tau,\nu) \le e^a \middle| \tau \ge \nu\right)\\
&\overset{(a)}{=} \mathbb{E}_\infty\left[\mathbb{1}_{\{\nu\le\tau<\nu+K_\gamma(1-\epsilon),\, \mathcal{L}(\tau,\nu)\le e^a\}}\mathcal{L}(\tau,\nu) \middle| \tau \ge \nu\right]\\
&\le e^a \mathbb{P}_\infty\big(\nu \le \tau < \nu + K_\gamma(1-\epsilon) | \tau \ge \nu\big)\\
&\overset{(b)}{\le} \frac{K_\gamma(1-\epsilon)e^a}{\gamma} \xrightarrow{\gamma\to\infty} 0,
\end{aligned} \tag{A.17}$$

where (a) follows by a change of measure argument; and (b) follows from the fact (see the proof of Theorem 1 in [22]) that for any positive integer $i < \gamma$, if $\mathbb{E}_\infty[\tau] \ge \gamma$, then there exists some $\nu \ge 1$ such that

$$\mathbb{P}_\infty(\tau \ge \nu) > 0, \text{ and } \mathbb{P}_\infty(\tau < \nu + i | \tau \ge \nu) \le \frac{i}{\gamma}. \tag{A.18}$$

Combining (A.12), (A.13), (A.16), (A.17), the fact that the upper bound in (A.17) is independent of τ, and the fact that for any stopping time τ, $\mathrm{WADD}(\tau) \ge \mathrm{CADD}(\tau)$, the theorem is established.

A.3 Proof of Theorem 2

Let $0 < \epsilon < J$, $\delta > 0$ and

$$n_b \triangleq \frac{b}{J-\epsilon}. \tag{A.19}$$

It can be shown that for any $\nu \ge 1$,

$$\begin{aligned}
&\operatorname{ess\,sup} \mathbb{E}_\nu\left[\frac{\tau_G(b)-\nu+1}{n_b} \middle| \tau_G(b) \ge \nu, \boldsymbol{X}[1,\nu-1]\right]\\
&\le \operatorname{ess\,sup} \sum_{\zeta=0}^{\infty} \mathbb{P}_\nu\left(\tau_G(b)-\nu+1 > \zeta n_b | \tau_G(b) \ge \nu, \boldsymbol{X}[1,\nu-1]\right)\\
&\le \sum_{\zeta=0}^{\infty} \operatorname{ess\,sup} \mathbb{P}_\nu\left(\tau_G(b) > \zeta n_b + \nu - 1 | \tau_G(b) \ge \nu, \boldsymbol{X}[1,\nu-1]\right)\\
&\le 1 + \sum_{\zeta=1}^{\infty} \operatorname{ess\,sup} \mathbb{P}_\nu\left(\tau_G(b) > \zeta n_b + \nu - 1 | \tau_G(b) \ge \nu, \boldsymbol{X}[1,\nu-1]\right).
\end{aligned} \tag{A.20}$$

For any $\zeta \geq 1$, it then follows that

$$
\begin{aligned}
&\mathbb{P}_\nu\left(\tau_G(b) > \zeta n_b + \nu - 1 | \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1]\right)\\
&= \mathbb{P}_\nu\left(\max_{1\leq k\leq \zeta n_b+\nu-1} W_G[k] < b \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1]\right)\\
&= \mathbb{P}_\nu\Bigg(\max_{1\leq k\leq \zeta n_b+\nu-1} \max_{k-\eta\leq i\leq k} \sum_{j=i}^{k} \log \frac{\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1])}{g(\boldsymbol{X}[j])} < b\\
&\qquad\qquad \Bigg| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg)\\
&\leq \mathbb{P}_\nu\Bigg(\bigcap_{r\in[\zeta]}\left\{\max_{rn_b+\nu-1-\eta\leq i\leq rn_b+\nu-1} \sum_{j=i}^{rn_b+\nu-1} \log \frac{\phi_i(\boldsymbol{X}[j]|\boldsymbol{X}[i,j-1])}{g(\boldsymbol{X}[j])} < b\right\}\\
&\qquad\qquad \Bigg| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg). \qquad \text{(A.21)}
\end{aligned}
$$

Without loss of generality, we choose η such that $\eta \geq n_b$ for large b. This further implies that $rn_b + \nu - \eta \leq (r-1)n_b + \nu$ for large b. As a result, for large b, (A.21) can be further upper bounded as follows:

$$
\begin{aligned}
&\mathbb{P}_\nu\left(\tau_G(b) > \zeta n_b + \nu - 1 | \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1]\right)\\
&\leq \mathbb{P}_\nu\left(\bigcap_{r\,\in\,[\zeta]} A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1]\right), \qquad \text{(A.22)}
\end{aligned}
$$

where for simplicity of notation, we denote the event

$$
A_r \triangleq \left\{\sum_{j=(r-1)n_b+\nu}^{rn_b+\nu-1} \log \frac{\phi_{(r-1)n_b+\nu}(\boldsymbol{X}[j]|\boldsymbol{X}[(r-1)n_b+\nu, j-1])}{g(\boldsymbol{X}[j])} < b\right\}, \qquad \text{(A.23)}
$$

for all $r \geq 1$. It is clear that $A_r \in \sigma(\boldsymbol{X}[(r-1)n_b + \nu, rn_b + \nu - 1])$. Then, it follows that

$$
\begin{aligned}
&\mathbb{P}_\nu\left(\bigcap_{r\,\in\,[\zeta]} A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1]\right)\\
&= \prod_{r=1}^{\zeta} \mathbb{P}_\nu\left(A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1], \bigcap_{j\,\in\,[r-1]} A_j\right). \qquad \text{(A.24)}
\end{aligned}
$$

This further implies that

$$\begin{aligned}
&\operatorname{ess\,sup} \mathbb{P}_\nu \left(\bigcap_{r \in [\zeta]} A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1] \right) \\
&\leq \prod_{r=1}^{\zeta} \operatorname{ess\,sup} \mathbb{P}_\nu \left(A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1], \bigcap_{j \in [r-1]} A_j \right). \qquad \text{(A.25)}
\end{aligned}$$

Then, if the following holds that for any $r \geq 1$,

$$\operatorname{ess\,sup} \mathbb{P}_\nu \left(A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1], \bigcap_{j \in [r-1]} A_j \right) \leq \delta, \qquad \text{(A.26)}$$

where δ is independent of ν, and can be arbitrarily small for large b, then

$$\operatorname{ess\,sup} \mathbb{P}_\nu \left(\bigcap_{r \in [\zeta]} A_r \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1] \right) \leq \delta^\zeta, \qquad \text{(A.27)}$$

which together with (A.20) implies that

$$\sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu \left[\frac{\tau_G(b) - \nu + 1}{n_b} \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \leq 1 + \sum_{\zeta=1}^{\infty} \delta^\zeta = \frac{1}{1-\delta}. \qquad \text{(A.28)}$$

This in turn implies that

$$\sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu \left[\tau_G(b) - \nu + 1 \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \leq \frac{b}{(J-\epsilon)(1-\delta)}. \qquad \text{(A.29)}$$

Since (A.29) holds for all ϵ small enough we have that

$$\sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu \left[\tau_G(b) - \nu + 1 \middle| \tau_G(b) \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \leq \frac{b}{J(1-\delta)}. \qquad \text{(A.30)}$$

Furthermore, since $\delta \to 0$ as $b \to \infty$, the proof is complete if we can show (A.26) is true.

In the following, we prove that (A.26) is true. We first note that according to our notation,

$$\begin{aligned}&\log \mathcal{L}(rn+\nu-1,(r-1)n+\nu)\\&=\sum_{j=(r-1)n+\nu}^{rn+\nu-1}\log\frac{\phi_{(r-1)n+\nu}(\boldsymbol{X}[j]|\boldsymbol{X}[(r-1)n+\nu,j-1])}{g(\boldsymbol{X}[j])}.\end{aligned}\tag{A.31}$$

By the Markov property of the problem model as in (2.4) and (2.5), it follows that

$$\begin{aligned}&\mathbb{P}_\nu\left(\frac{1}{n}\log\mathcal{L}(rn+\nu-1,(r-1)n+\nu)<\frac{J}{1+\epsilon}\middle|\tau_G(b)\geq\nu,\boldsymbol{X}[1,\nu-1],\bigcap_{j\in[r-1]}A_j\right)\\&=\sum_{\boldsymbol{E}\in\mathcal{E}}\mathbb{P}_\nu\left(\frac{1}{n}\log\mathcal{L}(rn+\nu-1,(r-1)n+\nu)<\frac{J}{1+\epsilon},\boldsymbol{S}[(r-1)n+\nu]=\boldsymbol{E}\right.\\&\qquad\qquad\left.\middle|\tau_G(b)\geq\nu,\boldsymbol{X}[1,\nu-1],\bigcap_{j\in[r-1]}A_j\right)\\&=\sum_{\boldsymbol{E}\in\mathcal{E}}\mathbb{P}_\nu\left(\frac{1}{n}\log\mathcal{L}(rn+\nu-1,(r-1)n+\nu)<\frac{J}{1+\epsilon}\middle|\boldsymbol{S}[(r-1)n+\nu]=\boldsymbol{E}\right)\\&\qquad\cdot\mathbb{P}_\nu\left(\boldsymbol{S}[(r-1)n+\nu]=\boldsymbol{E}\middle|\tau_G(b)\geq\nu,\boldsymbol{X}[1,\nu-1],\bigcap_{j\in[r-1]}A_j\right).\end{aligned}\tag{A.32}$$

From Lemma A.1. in [39], it follows that for any $\boldsymbol{E}\in\mathcal{E}$ and any $r\geq 1$,

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}_\nu\left(\frac{1}{n}\log\mathcal{L}(rn+\nu-1,(r-1)n+\nu)<\frac{J}{1+\epsilon}\middle|\boldsymbol{S}[(r-1)n+\nu]=\boldsymbol{E}\right)\\&=0.\end{aligned}\tag{A.33}$$

It then follows that for any $r \geq 1$,

$$\lim_{n\to\infty} \operatorname{ess\,sup} \mathbb{P}_\nu \Bigg(\sum_{j=(r-1)n+\nu}^{rn+\nu-1} \frac{1}{n} \log \frac{\phi_{(r-1)n+\nu}(\boldsymbol{X}[j]|\boldsymbol{X}[(r-1)n+\nu, j-1])}{g(\boldsymbol{X}[j])}$$
$$< \frac{J}{1+\epsilon} \Bigg| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1], \bigcap_{j\,\in\,[r-1]} A_j \Bigg)$$
$$= \lim_{n\to\infty} \operatorname{ess\,sup} \mathbb{P}_\nu \Bigg(\frac{1}{n} \log \mathcal{L}(rn+\nu-1, (r-1)n+\nu)$$
$$< \frac{J}{1+\epsilon} \Bigg| \tau_G(b) \geq \nu, \boldsymbol{X}[1,\nu-1], \bigcap_{j\,\in\,[r-1]} A_j \Bigg) = 0, \tag{A.34}$$

which further implies that (A.26) is true. This concludes the proof.

A.4 Proof of Lemma 2

Note that

$$\begin{aligned}
&\mathbb{E}_\infty[W_{SR}[k]|\boldsymbol{X}[1,k-1]] \\
&= \mathbb{E}_\infty \left[\sum_{\boldsymbol{E}\,\in\,\mathcal{E}} r_{\boldsymbol{E}}[k] \middle| \boldsymbol{X}[1,k-1] \right] \\
&= \mathbb{E}_\infty \left[\sum_{\boldsymbol{E}\,\in\,\mathcal{E}} \left(\left[\alpha_{\boldsymbol{E}} + \sum_{\boldsymbol{E}'\,\in\,\mathcal{E}} r_{\boldsymbol{E}'}[k-1]\lambda_{\boldsymbol{E}',\boldsymbol{E}} \right] \frac{p_{\boldsymbol{E}}(\boldsymbol{X}[k])}{g(\boldsymbol{X}[k])} \right) \middle| \boldsymbol{X}[1,k-1] \right] \\
&= 1 + \sum_{\boldsymbol{E}\,\in\,\mathcal{E}} \sum_{\boldsymbol{E}'\,\in\,\mathcal{E}} \lambda_{\boldsymbol{E}',\boldsymbol{E}} r_{\boldsymbol{E}'}[k-1] \\
&= 1 + \sum_{\boldsymbol{E}\,\in\,\mathcal{E}} r_{\boldsymbol{E}}[k-1] \\
&= 1 + W_{SR}[k-1],
\end{aligned} \tag{A.35}$$

which implies that $\{W_{SR}[k]-k\}_{k=1}^{\infty}$ is a martingale under $\mathbb{P}_\infty(\cdot)$. It can also be shown that $\mathbb{E}_\infty[W_{SR}[k]-k] = 0$. As a result, by the optimal stopping theorem (see e.g., [12]), it follows that $\mathbb{E}_\infty[W_{SR}[\tau_{SR}(b)] - \tau_{SR}(b)] = 0$. This further implies that

$$\mathbb{E}_\infty[\tau_{SR}(b)] = \mathbb{E}_\infty[W_{SR}[\tau_{SR}(b)]] \geq e^b. \tag{A.36}$$

A.5 Proof of Lemma 3

Let

$$\tau_S(b) \triangleq \inf\left\{k \geq 1 : W_S[k] \triangleq \sum_{j=1}^{k} \log \frac{\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1, j-1])}{g(\boldsymbol{X}[j])} \geq b\right\}. \tag{A.37}$$

By expressing the algorithm in (2.56) - (2.60) as a sequence of i.i.d. circles of (A.37) it can be easily shown that by using Wald's identity (see e.g., [11])

$$\mathbb{E}_\infty[\tau_{CE}] \geq \frac{\mathbb{E}_\infty[\tau_S(b)]}{\mathbb{P}_\infty(W_S[\tau_S(b)] \geq b)} \geq \frac{1}{\mathbb{P}_\infty(W_S[\tau_S(b)] \geq b)}. \tag{A.38}$$

Consider the event

$$A_i \triangleq \left\{\prod_{j=1}^{i} \frac{\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1], \ldots, \boldsymbol{X}[j-1])}{g(\boldsymbol{X}[j])} \geq e^b, \tau_S(b) = i\right\}. \tag{A.39}$$

It then follows that

$$\begin{aligned}
&\mathbb{P}_\infty(W_S[\tau_S(b)] \geq b) \\
&== \sum_{i=1}^{\infty} \mathbb{E}_\infty[\mathbb{1}_{\{A_i\}}] \\
&= \sum_{i=1}^{\infty} \mathbb{E}_\infty\left[\prod_{j=1}^{i} \frac{\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1, j-1])}{g(\boldsymbol{X}[j])} \prod_{j=1}^{i} \frac{g(\boldsymbol{X}[j])}{\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1, j-1])} \mathbb{1}_{\{A_i\}}\right] \\
&\leq e^{-b} \sum_{i=1}^{\infty} \mathbb{E}_\infty\left[\prod_{j=1}^{i} \frac{\phi_1(\boldsymbol{X}[j]|\boldsymbol{X}[1, j-1])}{g(\boldsymbol{X}[j])} \mathbb{1}_{\{A_i\}}\right] \\
&\leq e^{-b} \sum_{i=1}^{\infty} \mathbb{P}_1(A_i) \\
&\leq e^{-b}.
\end{aligned} \tag{A.40}$$

The result then follows by combining (A.38) and (A.40).

APPENDIX B

PROOFS FOR CHAPTER 3

B.1 Useful Lemmas

The proofs of Chapter 3 rely on the following lemmas:

Lemma 8. *For any stopping time τ and $N \geq 1$ define the truncated version of τ by $\tau^{(N)} \triangleq \min\{\tau, N\}$. We then have that*

$$\text{WADD}(\tau^{(N)}) \leq \text{WADD}(\tau). \tag{B.1}$$

Proof. Fix $\nu \geq 1$. Consider initially that $N \geq \nu$. Then, since $\{\tau^{(N)} \geq \nu\} = \{\min\{\tau, N\} \geq \nu\} = \{\tau \geq \nu\} \cap \{N \geq \nu\}$, we have that $\{\tau^{(N)} \geq \nu\} = \{\tau \geq \nu\}$. Since $\tau^{(N)} \leq \tau$, this implies that for any $N \geq \nu$ and any $\boldsymbol{S}$ we have that

$$\begin{aligned}
&\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&= \mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau^{(N)} - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\leq \mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\right].
\end{aligned} \tag{B.2}$$

For the case of $N < \nu$, we have that that $\mathbb{P}_{\nu}^{\boldsymbol{S}}(\tau^{(N)} \geq \nu) = 0$, which implies that by convention for any $N < \nu$ and any $\boldsymbol{S}$ we have that

$$\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] = 1. \tag{B.3}$$

Furthermore, note that for any $\boldsymbol{S}$ we have that

$$\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau - \nu + 1 | \tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \geq 1. \tag{B.4}$$

From (B.2) - (B.4) we have that for any $\nu \geq 1$ and any $\boldsymbol{S}$

$$\begin{aligned}&\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau^{(N)}-\nu+1|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right]\\&\leq\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau-\nu+1|\tau\geq\nu,\boldsymbol{X}[1,\nu-1]\right].\end{aligned}\tag{B.5}$$

By taking the sup and ess sup on both sides, with respect to the changepoint and history of observations respectively (4.4), the lemma is established. □

Lemma 9. *Let $C>0$, τ a stopping time adapted to $\boldsymbol{X}$, and $\Phi:\mathbb{R}\mapsto\mathbb{R}$ a function satisfying $|\Phi(x)|\leq C$ for all $x\in\mathbb{R}$. Then for any $\boldsymbol{\lambda}\in\mathcal{A}$ we have that*

$$\lim_{N\to\infty}\mathbb{E}_{\infty}\left[\sum_{k=0}^{\tau^{(N)}-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right]=\mathbb{E}_{\infty}\left[\sum_{k=0}^{\tau-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right].\tag{B.6}$$

Proof. Note that since $\tau\geq\tau^{(N)}$ we have that

$$\mathbb{E}_{\infty}\left[\sum_{k=0}^{\tau-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right]=\mathbb{E}_{\infty}\left[\sum_{k=0}^{\tau^{(N)}-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right]+\mathbb{E}_{\infty}\left[\sum_{k=\tau^{(N)}}^{\tau-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right].\tag{B.7}$$

Furthermore, note that by using Jensen's and triangle inequalities together with the assumption that $\Phi(x)$ is bounded we have that

$$\begin{aligned}\mathbb{E}_{\infty}\left[\sum_{k=\tau^{(N)}}^{\tau-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right]&\leq\left|\mathbb{E}_{\infty}\left[\sum_{k=\tau^{(N)}}^{\tau-1}\Phi(W_{\boldsymbol{\lambda}}[k])\right]\right|\\&\leq\mathbb{E}_{\infty}\left[\sum_{k=\tau^{(N)}}^{\tau-1}\left|\Phi(W_{\boldsymbol{\lambda}}[k])\right|\right]\\&\leq\mathbb{E}_{\infty}[\tau-\tau^{(N)}]\\&=\mathbb{E}_{\infty}[(\tau-N)^{+}].\end{aligned}\tag{B.8}$$

Since $(\tau - N)^+$ is a non-negative random variable, we then note that

$$\begin{aligned}\mathbb{E}_\infty[(\tau - N)^+] &= \sum_{j=0}^{\infty} \mathbb{P}_\infty((\tau - N)^+ > j) \\ &= \sum_{j=0}^{\infty} \mathbb{P}_\infty(\tau > j + N) \\ &= \sum_{j=N}^{\infty} \mathbb{P}_\infty(\tau > j), \end{aligned} \tag{B.9}$$

which since, by assumption

$$\mathbb{E}_\infty[\tau] = \sum_{j=0}^{\infty} \mathbb{P}_\infty(\tau > j) < \infty \tag{B.10}$$

implies that

$$\lim_{N\to\infty} \mathbb{E}_\infty[(\tau - N)^+] = \lim_{N\to\infty} \mathbb{P}_\infty(\tau > N) = 0. \tag{B.11}$$

As a result, from (B.8) we have that

$$\lim_{N\to\infty} \mathbb{E}_\infty \left[\sum_{k=\tau^{(N)}}^{\tau-1} \Phi_{\boldsymbol{\lambda}}(W[k]) \right] = 0. \tag{B.12}$$

After taking the limit in both sides of (B.7) and using eq. (B.12) the lemma is established. □

B.2 Proof of Theorem 5

Fix $\boldsymbol{\alpha} \in \mathcal{A}$. Due to the presence of the sup and ess sup in (3.4), we have that for any path $\boldsymbol{S}$, $\nu \geq 1$, stopping time τ and $N \geq 1$

$$\begin{aligned}
\mathrm{WADD}(\tau^{(N)}) &\geq \mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&= \mathbb{E}_\nu^{\boldsymbol{S}}\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\overset{\text{(a)}}{=} \mathbb{E}_\infty\left[\sum_{j=\nu}^{\infty} \Gamma_{\boldsymbol{S}}(j-1, \nu)\mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right]
\end{aligned} \tag{B.13}$$

where (a) follows by changing the measure to $\mathbb{P}_\infty(\cdot)$. By multiplying both sides of the inequality (B.13) with $\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu - 1])^+$ and taking the expected value under $\mathbb{E}_\infty[\cdot]$ we have that

$$\begin{aligned}
&\mathbb{E}_\infty\left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu - 1])^+ \mathrm{WADD}(\tau^{(N)})\right] \\
&\geq \mathbb{E}_\infty\Bigg[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu - 1])^+ \mathbb{E}_\infty\Bigg[\sum_{j=\nu}^{\infty} \Gamma_{\boldsymbol{S}}(j-1, \nu)\mathbb{1}_{\{\tau^{(N)} \geq j\}} \\
&\qquad\qquad \Bigg| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\Bigg]\Bigg] \\
&\overset{\text{(b)}}{=} \mathbb{E}_\infty\Bigg[\mathbb{E}_\infty\Bigg[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu - 1])^+ \sum_{j=\nu}^{\infty} \Gamma_{\boldsymbol{S}}(j-1, \nu)\mathbb{1}_{\{\tau^{(N)} \geq j\}} \\
&\qquad\qquad \Bigg| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\Bigg]\Bigg] \\
&\overset{\text{(c)}}{=} \mathbb{E}_\infty\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu - 1])^+ \Gamma_{\boldsymbol{S}}(j-1, \nu)\mathbb{1}_{\{\tau^{(N)} \geq j\}}\right],
\end{aligned} \tag{B.14}$$

where (b) follows since $\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}(1 - W_{\boldsymbol{\alpha}}[\nu - 1])^+$ is $\sigma(\boldsymbol{X}[1, \nu - 1])$-measurable and, hence, can go inside the expectation since the conditioning is with respect to $\boldsymbol{X}[1, \nu - 1]$; and (c) follows from the tower property of expectations. By summing on both sides of (B.14) over ν from $\nu = 1$ to $\nu = N$, and due to the linearity of expectation and the fact that $\tau^{(N)} \leq N$, we have that

$$\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} \mathbb{1}_{\{\tau^{(N)} \geq \nu\}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \mathrm{WADD}(\tau^{(N)}) \right]$$
$$\geq \mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} \sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} \geq \nu\}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \mathbb{1}_{\{\tau^{(N)} \geq j\}} \right], \quad \text{(B.15)}$$

which in turn implies that

$$\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \mathrm{WADD}(\tau^{(N)}) \right]$$
$$\geq \mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} \sum_{j=\nu}^{\tau^{(N)}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right]$$
$$\overset{\text{(d)}}{=} \mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right], \quad \text{(B.16)}$$

where (d) follows after changing the order of the summation. By the linearity of expectation, $\mathrm{WADD}(\tau^{(N)})$ can go outside of the expectation since it is a constant, hence we then have that

$$\mathrm{WADD}(\tau^{(N)}) \geq \frac{\mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right]}{\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \right]}. \quad \text{(B.17)}$$

By taking the sup with respect to $\boldsymbol{S}$, and since the right-hand side fraction depends on $\boldsymbol{S}$ only through $\boldsymbol{S}[1, N-1]$, we have that

$$\mathrm{WADD}(\tau^{(N)}) \geq \sup_{\boldsymbol{S}[1,N-1]} \frac{\mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right]}{\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \right]}. \quad \text{(B.18)}$$

Since the denominator in the right-hand side does not depend on $\boldsymbol{S}$, we have that

$$\mathrm{WADD}(\tau^{(N)}) \geq \frac{\sup_{\boldsymbol{S}[1,N-1]} \mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right]}{\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \right]}. \quad \text{(B.19)}$$

To proceed, we further bound the numerator in (B.19). For $1 \leq n \leq N-1$, define the following function:

$$\begin{aligned}
&\Psi_{n,N-1}(\boldsymbol{S}[1,n-1], \boldsymbol{S}[n+1,N-1]) \\
&\triangleq \sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty \left[\sum_{j=1}^{N} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \mathbb{1}_{\{\tau^{(N)} \geq j\}} \right]. \quad \text{(B.20)}
\end{aligned}$$

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

$$\begin{aligned}
&\sup_{\boldsymbol{S}[1,N-1]} \mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right] \\
&= \sup_{\boldsymbol{S}[1,n-1], \boldsymbol{S}[n+1,N-1]} \left[\sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty \left[\sum_{j=1}^{N} \sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \mathbb{1}_{\{\tau^{(N)} \geq j\}} \right] \right] \\
&= \sup_{\boldsymbol{S}[1,n-1], \boldsymbol{S}[n+1,N-1]} \Psi_{n,N-1}(\boldsymbol{S}[1,n-1], \boldsymbol{S}[n+1,N-1]). \quad \text{(B.21)}
\end{aligned}$$

Note that under $\mathbb{P}_\infty(\cdot)$ and for j such that $1 \leq j \leq n \leq N-1$ we have that

$$\sum_{\nu=1}^{j} (1 - W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \mathbb{1}_{\{\tau^{(N)} \geq j\}} \quad \text{(B.22)}$$

is independent of $\boldsymbol{S}[n]$. For $1 \leq n < j \leq N$ we have that

$$\begin{aligned}
&\sum_{\nu=1}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \\
&= \sum_{\nu=1}^{n}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \\
&\quad + \sum_{\nu=n+1}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}} \\
&= \Gamma_{\boldsymbol{S}}(n,n)\left(\sum_{\nu=1}^{n}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\left(\prod_{\substack{i=\nu\\ i\neq n}}^{j-1}\Gamma_{\boldsymbol{S}}(i,i)\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\right) \\
&\quad + \sum_{\nu=n+1}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}},
\end{aligned} \tag{B.23}$$

where under $\mathbb{P}_{\infty}(\cdot)$ the dependence from $\boldsymbol{S}[n]$ is only through the likelihood ratio $\Gamma_{\boldsymbol{S}}(n,n)$ of the first term.

For $1 \leq j \leq N$ and $1 \leq n \leq N-1$ define

$$A_{j,n} \triangleq \left(\sum_{\nu=1}^{n}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\left(\prod_{\substack{i=\nu\\ i\neq n}}^{j}\Gamma_{\boldsymbol{S}}(i,i)\right)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\right)\mathbb{1}_{\{j>n\}} \tag{B.24}$$

and

$$\begin{aligned}
B_{j,n} \triangleq &\left(\sum_{\nu=1}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\right)\mathbb{1}_{\{j\leq n\}} \\
&+ \left(\sum_{\nu=n+1}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}}\right)\mathbb{1}_{\{j>n\}}.
\end{aligned} \tag{B.25}$$

As a result, from eqs. (B.23) - (B.25) we have that for $1 \leq j \leq N$ and $1 \leq n \leq N-1$

$$\sum_{\nu=1}^{j}(1 - W_{\boldsymbol{\alpha}}[\nu-1])^{+}\Gamma_{\boldsymbol{S}}(j-1,\nu)\mathbb{1}_{\{\tau^{(N)}\geq j\}} = \Gamma_{\boldsymbol{S}}(n,n)A_{j,n} + B_{j,n}. \tag{B.26}$$

Then, from eqs. (B.20), (B.26) we have that

$$\begin{aligned}\Psi_{n,N-1}(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]) &= \sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty\left[\sum_{j=1}^{N}\left(\Gamma_{\boldsymbol{S}}(n,n)A_{j,n}+B_{j,n}\right)\right] \\ &= \sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty\left[\Gamma_{\boldsymbol{S}}(n,n)\sum_{j=1}^{N}A_{j,n}+\sum_{j=1}^{N}B_{j,n}\right].\end{aligned} \tag{B.27}$$

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 $\boldsymbol{E} \in \mathcal{E}$

$$\begin{aligned}&\sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty\left[\Gamma_{\boldsymbol{S}}(n,n)\sum_{j=1}^{N}A_{j,n}+\sum_{j=1}^{N}B_{j,n}\right] \\ &= \sup_{\boldsymbol{S}[n]} \mathbb{E}_\infty\left[\left(\prod_{\ell\in\boldsymbol{S}[n]}\frac{f_\ell(X_\ell[n])}{g_\ell(X_\ell[n])}\right)\sum_{j=1}^{N}A_{j,n}+\sum_{j=1}^{N}B_{j,n}\right] \\ &\geq \mathbb{E}_\infty\left[\left(\prod_{\ell\in\boldsymbol{E}}\frac{f_\ell(X_\ell[n])}{g_\ell(X_\ell[n])}\right)\sum_{j=1}^{N}A_{j,n}+\sum_{j=1}^{N}B_{j,n}\right],\end{aligned} \tag{B.28}$$

which together with eq. (B.27) implies that

$$\begin{aligned}&\Psi_{n,N-1}(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]) \\ &\geq \mathbb{E}_\infty\left[\left(\prod_{\ell\in\boldsymbol{E}}\frac{f_\ell(X_\ell[n])}{g_\ell(X_\ell[n])}\right)\sum_{j=1}^{N}A_{j,n}+\sum_{j=1}^{N}B_{j,n}\right].\end{aligned} \tag{B.29}$$

By averaging both sides of eq. (B.29) with respect to $\boldsymbol{\alpha}$ we then have that

$$\begin{aligned}
&\Psi_{n,N-1}(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]) \\
&= \sum_{\boldsymbol{E}\in\mathcal{E}} \alpha_{\boldsymbol{E}} \Psi_{n,N-1}(\boldsymbol{S}[1,n-1],\boldsymbol{S}[n+1,N-1]) \\
&\geq \sum_{\boldsymbol{E}\in\mathcal{E}} \boldsymbol{\alpha}_{\boldsymbol{E}} \mathbb{E}_\infty \left[\left(\prod_{\ell\in\boldsymbol{E}} \frac{f_\ell(X_\ell[n])}{g_\ell(X_\ell[n])} \right) \sum_{j=1}^{N} A_{j,n} + \sum_{j=}^{N} B_{j,n} \right] \\
&= \mathbb{E}_\infty \left[\left(\sum_{\boldsymbol{E}\in\mathcal{E}} \alpha_{\boldsymbol{E}} \left(\prod_{\ell\in\boldsymbol{E}} \frac{f_\ell(X_\ell[n])}{g_\ell(X_\ell[n])} \right) \right) \sum_{j=1}^{N} A_{j,n} + \sum_{j=1}^{N} B_{j,n} \right] \\
&= \mathbb{E}_\infty \left[\mathcal{L}_{\boldsymbol{\alpha}}(n,n) \left(\sum_{j=1}^{N} A_{j,n} \right) + \sum_{j=1}^{N} B_{j,n} \right] \\
&= \mathbb{E}_\infty \left[\sum_{j=1}^{N} \sum_{\nu=1}^{j} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \mathcal{L}_{\boldsymbol{\alpha}}(n,n) \left(\prod_{\substack{i=\nu \\ i\neq n}}^{j-1} \Gamma_{\boldsymbol{S}}(i,i) \right) \mathbb{1}_{\{\tau^{(N)}\geq j\}} \right]. \quad \text{(B.30)}
\end{aligned}$$

By unfolding eq. (B.21) in the same fashion with respect to all $1 \leq n \leq N-1$, it can be easily shown that

$$\begin{aligned}
&\sup_{\boldsymbol{S}[1,N-1]} \mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \Gamma_{\boldsymbol{S}}(j-1,\nu) \right] \\
&\geq \mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \mathcal{L}_{\boldsymbol{\alpha}}(j-1,\nu) \right], \quad \text{(B.31)}
\end{aligned}$$

which in turn together with (B.19) implies that

$$\begin{aligned}
\mathrm{WADD}(\tau^{(N)}) &\geq \frac{\mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \sum_{\nu=1}^{j} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \mathcal{L}_{\boldsymbol{\alpha}}(j-1,\nu) \right]}{\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \right]} \\
&= \frac{\mathbb{E}_\infty \left[\sum_{j=1}^{\tau^{(N)}} \left(\sum_{\nu=1}^{j-1} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \mathcal{L}_{\boldsymbol{\alpha}}(j-1,\nu) + (1-W_{\boldsymbol{\alpha}}[j-1])^+ \right) \right]}{\mathbb{E}_\infty \left[\sum_{\nu=1}^{\tau^{(N)}} (1-W_{\boldsymbol{\alpha}}[\nu-1])^+ \right]}. \quad \text{(B.32)}
\end{aligned}$$

From Lemma 1 of [19] we have that

$$\sum_{\nu=1}^{j-1}(1-W_{\boldsymbol{\alpha}}[\nu-1])^{+}\mathcal{L}_{\boldsymbol{\alpha}}(j-1,\nu)=W_{\boldsymbol{\alpha}}[j-1] \tag{B.33}$$

which together with (B.32) implies that

$$\begin{aligned}\mathrm{WADD}(\tau^{(N)}) &\geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=1}^{\tau^{(N)}}\left(W_{\boldsymbol{\alpha}}[j-1]+(1-W_{\boldsymbol{\alpha}}[j-1])^{+}\right)\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=1}^{\tau^{(N)}}(1-W_{\boldsymbol{\alpha}}[\nu-1])^{+}\right]}\\ &=\frac{\mathbb{E}_{\infty}\left[\sum_{j=1}^{\tau^{(N)}}\max\{W_{\boldsymbol{\alpha}}[j-1],1\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=1}^{\tau^{(N)}}(1-W_{\boldsymbol{\alpha}}[\nu-1])^{+}\right]}.\end{aligned} \tag{B.34}$$

Consider b chosen such that

$$\mathbb{E}_{\infty}[\tau_{M}(\boldsymbol{\alpha},b)]=\gamma. \tag{B.35}$$

Let $b'\geq b$ such that $b'>0$. Then, from Lemma 8 and (B.34) we have that

$$\begin{aligned}\mathrm{WADD}(\tau)\geq\mathrm{WADD}(\tau^{(N)}) &\geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=1}^{\tau^{(N)}}\max\{W_{\boldsymbol{\alpha}}[j-1],1\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=1}^{\tau^{(N)}}(1-W_{\boldsymbol{\alpha}}[\nu-1])^{+}\right]}\\ &\geq\frac{\mathbb{E}_{\infty}\left[\sum_{j=1}^{\tau^{(N)}}\min\{\max\{W_{\boldsymbol{\alpha}}[j-1],1\},e^{b'}\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=1}^{\tau^{(N)}}(1-W_{\boldsymbol{\alpha}}[\nu-1])^{+}\right]}.\end{aligned} \tag{B.36}$$

Note that

$$\left|\min\left\{\max\{W_{\boldsymbol{\alpha}}[j-1],1\},e^{b'}\right\}\right|\leq e^{b'} \tag{B.37}$$

and that since $W_{\boldsymbol{\alpha}}[j-1] \geq 0$

$$|(1 - W_{\boldsymbol{\alpha}}[j-1])^{+}| \leq 1. \tag{B.38}$$

Furthermore, since $\mathbb{E}_{\infty}[\tau] < \infty$ by assumption, by using Lemma 9 after taking the limit on both sides of (B.36) and doing a change of variables we have that

$$\mathrm{WADD}(\tau) \geq \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}. \tag{B.39}$$

Since (B.39) holds for arbitrary τ, we have that for any $\gamma > 0$

$$\begin{aligned}
\inf_{\tau \in C_{\gamma}} \mathrm{WADD}(\tau) &\geq \inf_{\tau \in C_{\gamma}} \frac{\mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\}\right]}{\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^{+}\right]} \\
&\geq \frac{\inf_{\tau \in C_{\gamma}} \mathbb{E}_{\infty}\left[\sum_{j=0}^{\tau-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\}\right]}{\sup_{\tau \in C_{\gamma}} \mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^{+}\right]}.
\end{aligned} \tag{B.40}$$

Note that the function $Q(x) \triangleq (1-x)^{+}$ is continuous and non-increasing with $Q(0) = 1$. As a result, from Theorem 1 of [19] we have that

$$\mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau_M(\boldsymbol{\alpha}, b)-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^{+}\right] = \sup_{\tau \in C_{\gamma}} \mathbb{E}_{\infty}\left[\sum_{\nu=0}^{\tau-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^{+}\right]. \tag{B.41}$$

Furthermore, note that the function $U(x) \triangleq -\min\left\{\max\{x, 1\}, e^{b'}\right\}$ is continuous and non-increasing in x with $U(0) = -\min\left\{1, e^{b'}\right\}$. As a result, from Theorem 1 of [19] we also have that

$$\begin{aligned}
&\inf_{\tau \in C_\gamma} \mathbb{E}_\infty \left[\sum_{j=0}^{\tau-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\} \right] \\
&= -\sup_{\tau \in C_\gamma} \mathbb{E}_\infty \left[-\sum_{j=0}^{\tau-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\} \right] \\
&= -\mathbb{E}_\infty \left[-\sum_{j=0}^{\tau_M(\boldsymbol{\alpha},b)-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\} \right] \\
&= \mathbb{E}_\infty \left[\sum_{j=0}^{\tau_M(\boldsymbol{\alpha},b)-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\} \right]. \qquad \text{(B.42)}
\end{aligned}$$

Then, from (B.40) - (B.42) we have that

$$\begin{aligned}
\inf_{\tau \in C_\gamma} \text{WADD}(\tau) &\geq \frac{\mathbb{E}_\infty \left[\sum_{j=0}^{\tau_M(\boldsymbol{\alpha},b)-1} \min\{\max\{W_{\boldsymbol{\alpha}}[j], 1\}, e^{b'}\} \right]}{\mathbb{E}_\infty \left[\sum_{\nu=0}^{\tau_M(\boldsymbol{\alpha},b)-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^+ \right]} \\
&\overset{\text{(e)}}{=} \frac{\mathbb{E}_\infty \left[\sum_{j=0}^{\tau_M(\boldsymbol{\alpha},b)-1} \max\{W_{\boldsymbol{\alpha}}[j], 1\} \right]}{\mathbb{E}_\infty \left[\sum_{\nu=0}^{\tau_M(\boldsymbol{\alpha},b)-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^+ \right]}, \qquad \text{(B.43)}
\end{aligned}$$

where (e) is implied since $W_{\boldsymbol{\alpha}}[j] < e^b \leq e^{b'}$ for $0 \leq j \leq \tau_M(\boldsymbol{\alpha}, b) - 1$ and since $b' > 0$. Furthermore, note that from the optimality of the CUSUM test for the classic QCD problem [19] we have that

$$\frac{\mathbb{E}_\infty \left[\sum_{j=0}^{\tau_M(\boldsymbol{\alpha},b)-1} \max\{W_{\boldsymbol{\alpha}}[j], 1\} \right]}{\mathbb{E}_\infty \left[\sum_{\nu=0}^{\tau_M(\boldsymbol{\alpha},b)-1} (1 - W_{\boldsymbol{\alpha}}[\nu])^+ \right]} = \overline{\text{WADD}}(\tau_M(\boldsymbol{\alpha}, b)). \qquad \text{(B.44)}$$

As a result, from (B.43) and (B.44) and since

$$\text{WADD}(\tau_M(\boldsymbol{\alpha}, b)) \geq \inf_{\tau \in C_\gamma} \text{WADD}(\tau) \qquad \text{(B.45)}$$

the theorem is established.

B.3 Proof of Lemma 5

Fix $\boldsymbol{\alpha} \in \mathcal{A}$, $b > 0$ and $N \geq 1$. For purposes of presentation of this proof, we denote the stopping $\tau_M(\boldsymbol{\lambda}_U, b)$ with uniform weights and threshold b by simply τ_M 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 $\tau_M^{(N)} \triangleq \min\{\tau_M, N\}$. Note that by employing a change of measure similar to the one in (B.13) we have that for any $\nu \geq 1$ and any $\boldsymbol{S}$

$$\begin{aligned}
V_\nu &\triangleq \mathbb{E}_\nu^{\boldsymbol{S}}\left[\tau_M^{(N)} - \nu + 1 \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right] \\
&= \mathbb{E}_\nu^{\boldsymbol{S}}\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\left\{\tau_M^{(N)} \geq j\right\}} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right] \\
&= \mathbb{E}_\nu^{\boldsymbol{S}}\left[\sum_{j=\nu}^{N} \mathbb{1}_{\left\{\tau_M^{(N)} \geq j\right\}} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right] \\
&= \mathbb{E}_\infty\left[\sum_{j=\nu}^{N} \Gamma_{\boldsymbol{S}}(j-1,\nu) \mathbb{1}_{\left\{\tau_M^{(N)} \geq j\right\}} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right] \\
&= 1 + \mathbb{E}_\infty\left[\sum_{j=\nu+1}^{N} \Gamma_{\boldsymbol{S}}(j-1,\nu) \mathbb{1}_{\left\{\tau_M^{(N)} \geq j\right\}} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right] \\
&\overset{\text{(a)}}{=} 1 + \mathbb{E}_\infty\left[\sum_{j=\nu+1}^{N} \Gamma_{\boldsymbol{S}}(j-1,\nu) \left(\prod_{i=\nu}^{j-1} \mathbb{1}_{\{W[i] < e^b\}}\right) \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right],
\end{aligned} \tag{B.46}$$

where (a) follows since for $\nu + 1 \leq j \leq N$ we have that conditioned on $\left\{\tau_M^{(N)} \geq \nu\right\}$

$$\left\{\tau_M^{(N)} \geq j\right\} = \bigcap_{i=\nu}^{j-1} \{W[i] < e^b\}. \tag{B.47}$$

To proceed, we establish that for any $1 \leq \nu \leq N-1$

$$V_\nu = 1 + \mathbb{E}_\infty\left[\Gamma_{\boldsymbol{S}}(\nu,\nu) \mathbb{1}_{\{W[\nu] < e^b\}} V_{\nu+1} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\right], \tag{B.48}$$

with $V_\nu = 1$ for all $\nu \geq N$. First of all, from the definition of V_ν we have that

$$\begin{aligned} V_N &= \mathbb{E}_N^{\boldsymbol{S}}\left[\tau_M^{(N)} - N + 1 \middle| \tau_M^{(N)} \geq N, \boldsymbol{X}[1, N-1]\right] \\ &= \mathbb{E}_N^{\boldsymbol{S}}\left[N - N + 1 \middle| \tau_M^{(N)} \geq N, \boldsymbol{X}[1, N-1]\right] \\ &= 1. \end{aligned} \tag{B.49}$$

In addition, for $\nu \geq N+1$ the event $\{\tau_M^{(N)} \geq \nu\}$ cannot occur, hence, we have that $V_\nu = 1$ for all $\nu \geq N$. Furthermore, note that $\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu]<e^b\}}$ is present in all terms of the summation in (B.46), hence

$$\begin{aligned} V_\nu = 1 + \mathbb{E}_\infty\Bigg[&\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu]<e^b\}} \\ &\cdot \sum_{j=\nu+1}^{N} \Gamma_{\boldsymbol{S}}(j-1,\nu+1)\left(\prod_{i=\nu+1}^{j-1} \mathbb{1}_{\{W[i]<e^b\}}\right) \\ &\Bigg| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg] \\ = 1 + \mathbb{E}_\infty\Bigg[&\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu+1]<e^b\}} \\ &\cdot \left(1 + \sum_{j=\nu+2}^{N-1} \Gamma_{\boldsymbol{S}}(j-1,\nu+1)\left(\prod_{i=\nu+1}^{j-1} \mathbb{1}_{\{W[i]<e^b\}}\right)\right) \\ &\Bigg| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg] \\ \overset{(b)}{=} 1 + \mathbb{E}_\infty\Bigg[&\mathbb{E}_\infty\Bigg[\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu]<e^b\}} \\ &\cdot \left(1 + \sum_{j=\nu+2}^{N} \Gamma_{\boldsymbol{S}}(j-1,\nu+1)\left(\prod_{i=\nu+1}^{j-1} \mathbb{1}_{\{W[i]<e^b\}}\right)\right) \\ &\Bigg| \tau_M^{(N)} \geq \nu+1, \boldsymbol{X}[1,\nu]\Bigg]\Bigg| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg] \\ \overset{(c)}{=} 1 + \mathbb{E}_\infty\Bigg[&\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu]<e^b\}} \\ &\cdot \left(1 + \mathbb{E}_\infty\Bigg[\sum_{j=\nu+2}^{N} \Gamma_{\boldsymbol{S}}(j-1,\nu+1)\left(\prod_{i=\nu+1}^{j-1} \mathbb{1}_{\{W[i]<e^b\}}\right)\right. \\ &\left.\Bigg| \tau_M^{(N)} \geq \nu+1, \boldsymbol{X}[1,\nu]\Bigg]\right)\Bigg| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg] \\ \overset{(d)}{=} 1 + \mathbb{E}_\infty\Bigg[&\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu]<e^b\}} V_{\nu+1} \Bigg| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1,\nu-1]\Bigg], \end{aligned} \tag{B.50}$$

where (b) follows from the tower property of expectations; (c) follows since $\Gamma_{\boldsymbol{S}}(\nu,\nu)\mathbb{1}_{\{W[\nu]<e^b\}}$ is $\sigma(\boldsymbol{X}[1,\nu])$-measurable; and hence can go out of the conditional expectation, and (d) follows from (B.46).

We will now establish that V_ν is independent of $\boldsymbol{S}$ for all $\nu \geq 1$ and that it is a function of $\boldsymbol{X}[1,\nu]$ only through $W[\nu]$. First of all, we have already established that for $\nu \geq N$, $V_\nu = 1$, hence we only have to investigate the case of $\nu \leq N-1$. For $\nu \leq N-1$ since $\tau_M^{(N)}$ is truncated by N and since $\boldsymbol{X}[1,\nu-1]$ are independent from $\boldsymbol{S}$ we have to show that V_ν is independent of $\boldsymbol{S}[\nu,N]$ and that V_ν is a function of $\boldsymbol{X}[1,\nu-1]$ only through $W[\nu-1]$. For $1 \leq k \leq N-2$, assume that the statement holds for $V_{N-k}(W[N-1-k])$. From (B.48) we have that

$$\begin{aligned}
V_{N-(k+1)} &= 1 + \mathbb{E}_\infty\Bigg[\Gamma_{\boldsymbol{S}}(N-1-k, N-1-k) \\
&\qquad \cdot \mathbb{1}_{\{W[N-1-k]<e^b\}} V_{N-k}(W[N-1-k]) \\
&\qquad \Bigg| \tau_M^{(N)} \geq N-1-k, \boldsymbol{X}[1,N-2-k]\Bigg] \\
&\overset{\text{(e)}}{=} 1 + \mathbb{E}_\infty\Bigg[\Gamma_{\boldsymbol{S}}(N-1-k, N-1-k)\mathbb{1}_{\{\max\{W[N-2-k],1\}\mathcal{L}(N-1-k,N-1-k)<e^b\}} \\
&\qquad \cdot V_{N-k}(\max\{W[N-2-k],1\}\mathcal{L}(N-1-k,N-1-k)) \\
&\qquad \Bigg| \tau_M^{(N)} \geq N-1-k, \boldsymbol{X}[1,N-2-k]\Bigg] \\
&\overset{\text{(f)}}{=} 1 + \mathbb{E}_\infty\Bigg[\left(\prod_{\ell \in \boldsymbol{S}[N-1-k]} \frac{f(X_\ell[N-1-k])}{g(X_\ell[N-1-k])}\right) \\
&\qquad \cdot \mathbb{1}_{\left\{\max\{W[N-2-k],1\}\left(\prod_{\ell \in \boldsymbol{S}[N-1-k]} \frac{f(X_\ell[N-1-k])}{g(X_\ell[N-1-k])}\right)<e^b\right\}} \\
&\qquad \cdot V_{N-k}\left(\max\{W[N-2-k],1\}\left(\prod_{\ell \in \boldsymbol{S}[N-1-k]} \frac{f(X_\ell[N-1-k])}{g(X_\ell[N-1-k])}\right)\right) \\
&\qquad \Bigg| \tau_M^{(N)} \geq N-1-k, \boldsymbol{X}[1,N-2-k]\Bigg],
\end{aligned} \tag{B.51}$$

where (e) follows from eq. (3.14); and (f) follows from (3.9). Note that, under $\mathbb{P}_\infty(\cdot)$, the distribution of the likelihood ratio in (B.51) is independent of $\boldsymbol{S}[N-1-k]$. As a result, we have that for all $\boldsymbol{E} \in \mathcal{E}$

$$
\begin{aligned}
V_{N-(k+1)} =& 1 + \mathbb{E}_\infty \Bigg[\left(\prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[N-1-k])}{g(X_\ell[N-1-k])} \right) \\
& \cdot \mathbb{1}_{\left\{ \max\{W[N-2-k],1\} \left(\prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[N-1-k])}{g(X_\ell[N-1-k])} \right) < e^b \right\}} \\
& \cdot V_{N-k} \left(\max\{W[N-2-k],1\} \left(\prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[N-1-k])}{g(X_\ell[N-1-k])} \right) \right) \\
& \Bigg| \tau_M^{(N)} \geq N-1-k, \boldsymbol{X}[1, N-2-k] \Bigg]. \qquad \text{(B.52)}
\end{aligned}
$$

From (B.52), we can then easily see that $V_{N-(k+1)}$ is independent of $\boldsymbol{S}$. Furthermore, since the likelihood ratio in (B.52) is independent of $\boldsymbol{X}[1, N-2-k]$ we have that $V_{N-(k+1)}$ is a function of $\boldsymbol{X}[1, N-2-k]$ only through $W[N-2-k]$. As a result, by induction we have that for all $\nu \geq 1$, V_ν is independent of $\boldsymbol{S}$ and depends on $\boldsymbol{X}[1, \nu-1]$ only through $W[\nu-1]$.

Following, note that for $\nu \geq 1$, from the independence of V_ν from $\boldsymbol{S}$ and eq. (B.48) we have that for all $\boldsymbol{E} \in \mathcal{E}$

$$
V_\nu = 1 + \mathbb{E}_\infty \left[\left(\prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[\nu])}{g(X_\ell[\nu])} \right) \mathbb{1}_{\{W[\nu] < e^b\}} V_{\nu+1} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right]. \qquad \text{(B.53)}
$$

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

$$
\begin{aligned}
V_\nu &= 1 + \sum_{\boldsymbol{E} \in \mathcal{E}} \alpha_{\boldsymbol{E}} \mathbb{E}_\infty \left[\left(\prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[\nu])}{g(X_\ell[\nu])} \right) \mathbb{1}_{\{W[\nu] < e^b\}} V_{\nu+1} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&= 1 + \mathbb{E}_\infty \left[\left(\sum_{\boldsymbol{E} \in \mathcal{E}} \alpha_{\boldsymbol{E}} \prod_{\ell \in \boldsymbol{E}} \frac{f(X_\ell[\nu])}{g(X_\ell[\nu])} \right) \mathbb{1}_{\{W[\nu] < e^b\}} V_{\nu+1} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&= 1 + \mathbb{E}_\infty \left[\mathcal{L}(\nu, \nu) \mathbb{1}_{\{W[\nu] < e^b\}} V_{\nu+1} \middle| \tau_M^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right]. \qquad \text{(B.54)}
\end{aligned}
$$

By unfolding the recursion in (B.54), it can be easily seen that for any $\nu \geq 1$

and $\boldsymbol{S}$

$$\begin{aligned}&\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau_M^{(N)}-\nu+1|\tau_M^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right]\\&=\overline{\mathbb{E}}_{\nu}^{\boldsymbol{\alpha}}\left[\tau_M^{(N)}-\nu+1|\tau_M^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right].\end{aligned}\tag{B.55}$$

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

$$\begin{aligned}&\lim_{N\to\infty}\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau_M^{(N)}-\nu+1\middle|\tau_M^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right]\\&=\lim_{N\to\infty}\frac{\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[(\tau_M^{(N)}-\nu+1)\mathbb{1}_{\{\tau_M^{(N)}\geq\nu\}}\middle|\boldsymbol{X}[1,\nu-1]\right]}{\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\mathbb{1}_{\left\{\tau_M^{(N)}\geq\nu\right\}}\middle|\boldsymbol{X}[1,\nu-1]\right]}\\&=\frac{\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\lim_{N\to\infty}(\tau_M^{(N)}-\nu+1)\mathbb{1}_{\{\tau_M^{(N)}\geq\nu\}}\middle|\boldsymbol{X}[1,\nu-1]\right]}{\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\lim_{N\to\infty}\mathbb{1}_{\{\tau_M^{(N)}\geq\nu\}}\middle|\boldsymbol{X}[1,\nu-1]\right]}\\&=\frac{\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[(\tau_M-\nu+1)\mathbb{1}_{\{\tau_M\geq\nu\}}|\boldsymbol{X}[1,\nu-1]\right]}{\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\mathbb{1}_{\{\tau_M\geq\nu\}}|\boldsymbol{X}[1,\nu-1]\right]}\\&=\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau_M-\nu+1|\tau_M\geq\nu,\boldsymbol{X}[1,\nu-1]\right].\end{aligned}\tag{B.56}$$

Similarly, it can be shown that

$$\begin{aligned}&\lim_{N\to\infty}\overline{\mathbb{E}}_{\nu}^{\boldsymbol{\alpha}}\left[\tau_M^{(N)}-\nu+1|\tau_M^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right]\\&=\overline{\mathbb{E}}_{\nu}^{\boldsymbol{\alpha}}\left[\tau_M-\nu+1|\tau_M\geq\nu,\boldsymbol{X}[1,\nu-1]\right].\end{aligned}\tag{B.57}$$

As a result, by taking the limit on both sides of (B.55) and using eqs. (B.56) and (B.57) we have that for all $\nu\geq 1$, $\boldsymbol{S}$

$$\mathbb{E}_{\nu}^{\boldsymbol{S}}\left[\tau_M-\nu+1|\tau_M\geq\nu,\boldsymbol{X}[1,\nu-1]\right]=\overline{\mathbb{E}}_{\nu}^{\boldsymbol{\alpha}}\left[\tau_M-\nu+1|\tau_M\geq\nu,\boldsymbol{X}[1,\nu-1]\right]\tag{B.58}$$

which in turn implies

$$\text{WADD}(\tau_M)=\overline{\text{WADD}}_{\boldsymbol{\alpha}}(\tau_M).\tag{B.59}$$

B.4 Proof of Lemma 6

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

$$\begin{aligned} \inf_{\boldsymbol{\beta}} \quad & q(\boldsymbol{\beta}) \\ \text{s.t.} \quad & \beta_{\boldsymbol{E}_j} \geq 0, \ \forall\, j \in [|\mathcal{E}|-1] \\ & \sum_{j=1}^{|\mathcal{E}|-1} \beta_{\boldsymbol{E}_j} \leq 1, \end{aligned} \tag{B.60}$$

where

$$\begin{aligned} q(\boldsymbol{\beta}) \triangleq \int_{\mathbb{R}^L} & \left((1 - \|\boldsymbol{\beta}\|_1)\, p_{\boldsymbol{E}_{|\mathcal{E}|}}(\boldsymbol{x}) + \sum_{j=1}^{|\mathcal{E}|-1} \beta_{\boldsymbol{E}_j} p_{\boldsymbol{E}_j}(\boldsymbol{x}) \right) \\ & \cdot \log \left(\frac{\left((1 - \|\boldsymbol{\beta}\|_1)\, p_{\boldsymbol{E}_{|\mathcal{E}|}}(\boldsymbol{x}) + \sum_{j=1}^{|\mathcal{E}|-1} \beta_{\boldsymbol{E}_j} p_{\boldsymbol{E}_j}(\boldsymbol{x}) \right)}{g(\boldsymbol{x})} \right) d\mu(\boldsymbol{x}). \end{aligned} \tag{B.61}$$

Denote by $\overset{*}{\boldsymbol{\beta}}$ the solution to (B.60). Then, the derivative at $\overset{*}{\boldsymbol{\beta}}$ is given by

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\overset{*}{\boldsymbol{\beta}}} = \mathbb{E}_{p_{\boldsymbol{E}_i}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] - \mathbb{E}_{p_{\boldsymbol{E}_{|\mathcal{E}|}}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right]. \tag{B.62}$$

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

Assume that $\overset{*}{\boldsymbol{\beta}}$ is a corner point. In this case we have that for all $i \in [|\mathcal{E}|-1]$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\overset{*}{\boldsymbol{\beta}}} = \sum_{\ell \,\in\, \boldsymbol{E}_{|\mathcal{E}|}} \left(D(f_\ell \| g_\ell) \mathbb{1}_{\{\ell \,\in\, \boldsymbol{E}_i\}} - D(g_\ell \| f_\ell) \mathbb{1}_{\{\ell \,\notin\, \boldsymbol{E}_i\}} \right) - \sum_{\ell \,\in\, \boldsymbol{E}_{|\mathcal{E}|}} D(f_i \| g_i) < 0, \tag{B.63}$$

which is a contradiction since

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\overset{*}{\boldsymbol{\beta}}} \geq 0 \tag{B.64}$$

must hold for all $i \in [|\mathcal{E}|-1]$ due to the fact that $\mathring{\boldsymbol{\beta}}$ is a minimum.

As a result, $\mathring{\boldsymbol{\beta}}$ is not a corner point, hence $\mathring{\boldsymbol{\beta}} = \left[\mathring{\beta}_{\boldsymbol{E}_1}, \ldots, \mathring{\beta}_{\boldsymbol{E}_l}, \ldots, 0\right]^\top$. In this case, for all $i \in [l]$ we have that

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\mathring{\boldsymbol{\beta}}} = 0, \tag{B.65}$$

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

$$\mathbb{E}_{p_{\boldsymbol{E}_i}}\left[\log\left(\frac{\overline{p}_{\mathring{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] = \mathbb{E}_{p_{\boldsymbol{E}_{|\mathcal{E}|}}}\left[\log\left(\frac{\overline{p}_{\mathring{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] \triangleq I'. \tag{B.66}$$

Furthermore, we have that since $\mathring{\beta}_{\boldsymbol{E}_j} = 0$ for $l+1 \leq j \leq |\mathcal{E}|-1$

$$\begin{aligned} I' &= \left(\sum_{j=1}^{l} \mathring{\beta}_{\boldsymbol{E}_j} + \left(1 - \sum_{j=1}^{l} \mathring{\beta}_{\boldsymbol{E}_j}\right)\right) I' \\ &= \left(\sum_{j=1}^{l} \mathring{\alpha}_{\boldsymbol{E}_j} + \mathring{\alpha}_{\boldsymbol{E}_{|\mathcal{E}|}}\right) I' \\ &= \sum_{j=1}^{|\mathcal{E}|} \mathring{\alpha}_{\boldsymbol{E}_j} \mathbb{E}_{p_{\boldsymbol{E}_j}}\left[\log\left(\frac{\overline{p}_{\mathring{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] \\ &= \mathbb{E}_{\overline{p}_{\mathring{\boldsymbol{\alpha}}}}\left[\log\left(\frac{\overline{p}_{\mathring{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] \\ &= \mathring{I} \\ &> 0. \end{aligned} \tag{B.67}$$

In addition, we have that for $l+1 \leq i \leq |\mathcal{E}|-1$

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{\boldsymbol{E}_i}}\right|_{\mathring{\boldsymbol{\beta}}} > 0. \tag{B.68}$$

This implies that for all $i \in [l] \cup \{|\mathcal{E}|\}$ and $l+1 \leq j \leq |\mathcal{E}|-1$

$$\mathbb{E}_{p_{\boldsymbol{E}_j}}\left[\log\left(\frac{\overline{p}_{\mathring{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] > \mathbb{E}_{p_{\boldsymbol{E}_i}}\left[\log\left(\frac{\overline{p}_{\mathring{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] = \mathring{I}. \tag{B.69}$$

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 $l+1 \leq i \leq L-1$, we then have that

$$\begin{aligned}
\mathbb{E}_{p_{E_i}}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] &= \mathbb{E}_{p_i}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] \\
&= \mathbb{E}_{p_i}\left[\log\left(\sum_{j=1}^{l}\overset{*}{\alpha}_j\frac{f_j(X_j[1])}{g_j(X_j[1])} + \overset{*}{\alpha}_L\frac{f_L(X_L[1])}{g_L(X_L[1])}\right)\right] \\
&= \mathbb{E}_{g}\left[\log\left(\sum_{j=1}^{l}\overset{*}{\alpha}_j\frac{f_j(X_j[1])}{g_j(X_j[1])} + \overset{*}{\alpha}_L\frac{f_L(X_L[1])}{g_L(X_L[1])}\right)\right] \\
&= \mathbb{E}_{g}\left[\log\left(\frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[1])}{g(\boldsymbol{X}[1])}\right)\right] \\
&< 0. \qquad \text{(B.70)}
\end{aligned}$$

We then have that from eqs. (B.62), (B.66), (B.67) and (B.70)

$$\left.\frac{\partial q(\boldsymbol{\beta})}{\partial \beta_{E_i}}\right|_{\overset{*}{\boldsymbol{\beta}}} < 0 \qquad \text{(B.71)}$$

for all $l+1 \leq i \leq L-1$, which leads to a contradiction, since (B.71) cannot hold at the minimum.

B.5 Proof of Theorem 8

Our upper bound analysis is based on the proof technique in [22]. Due to the structure of the M-CUSUM test described in (3.12) - (3.14), we have that for any $b > 0$

$$\mathrm{WADD}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)) = \sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}}[\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)]. \qquad \text{(B.72)}$$

Let $0 < \epsilon < \overset{*}{I}$ and

$$n_b \triangleq \frac{b}{\overset{*}{I} - \epsilon}. \qquad \text{(B.73)}$$

We then have that

$$\begin{aligned}
\sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}}\left[\frac{\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b}\right] &\overset{\text{(a)}}{=} \sup_{\boldsymbol{S}} \int_0^\infty \mathbb{P}_1^{\boldsymbol{S}}\left(\frac{\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b} > x\right) dx \\
&\overset{\text{(b)}}{\leq} \sup_{\boldsymbol{S}} \sum_{\zeta=0}^{\infty} \mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&= 1 + \sup_{\boldsymbol{S}} \sum_{\zeta=1}^{\infty} \mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b), \qquad \text{(B.74)}
\end{aligned}$$

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 (3.7) for $\boldsymbol{\alpha} = \overset{*}{\boldsymbol{\alpha}}$ by

$$Z[j] \triangleq \log \frac{\overline{p}_{\overset{*}{\boldsymbol{\alpha}}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}. \qquad \text{(B.75)}$$

For any path $\boldsymbol{S}$, $\zeta \geq 1$, we then have that

$$
\begin{aligned}
&\mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&= \mathbb{P}_1^{\boldsymbol{S}}\left(\max_{1\leq k\leq \zeta n_b} W_{\overset{*}{\boldsymbol{\alpha}}}[k] < e^b\right) \\
&\overset{(c)}{=} \mathbb{P}_1^{\boldsymbol{S}}\left(\max_{1\leq k\leq \zeta n_b} \max_{1\leq i\leq k} \mathcal{L}_{\overset{*}{\boldsymbol{\alpha}}}(k,i) < e^b\right) \\
&\overset{(d)}{=} \mathbb{P}_1^{\boldsymbol{S}}\left(\max_{1\leq k\leq \zeta n_b} \max_{1\leq i\leq k} \sum_{j=i}^{k} Z[j] < b\right) \\
&\overset{(e)}{\leq} \mathbb{P}_1^{\boldsymbol{S}}\left(\max_{1\leq i\leq r n_b} \sum_{j=i}^{r n_b} Z[j] < b,\ \forall\, r \in [\zeta]\right) \\
&\overset{(f)}{\leq} \mathbb{P}_1^{\boldsymbol{S}}\left(\sum_{j=(r-1)n_b+1}^{r n_b} Z[j] < b,\ \forall\, r \in [\zeta]\right) \\
&\overset{(g)}{=} \mathbb{P}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(r-1)n_b+1}^{r n_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon,\ \forall\, r \in [\zeta]\right) \\
&\overset{(h)}{=} \prod_{r=1}^{\zeta} \mathbb{P}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=(r-1)n_b+1}^{r n_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon\right),
\end{aligned}
\tag{B.76}
$$

where (c) follows from the definition of the M-CUSUM statistic (eq. (3.12)); (d) follows by taking the logarithm at both sides of the inequality; (e) and (f) by using the binning technique in [22]; (g) by diving both sides by n_b; and (h) by the independence of the observations over time.

Note that for $b > 0$ we then have that from (B.76)

$$
\begin{aligned}
&\sup_{\boldsymbol{S}} \sum_{\zeta=1}^{\infty} \mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&= \sup_{\boldsymbol{S}} \lim_{\xi \to \infty} \sum_{\zeta=1}^{\xi} \mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&\leq \lim_{\xi \to \infty} \sup_{\boldsymbol{S}} \sum_{\zeta=1}^{\xi} \mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&\leq \lim_{\xi \to \infty} \sum_{\zeta=1}^{\xi} \sup_{\boldsymbol{S}} \mathbb{P}_1^{\boldsymbol{S}}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&\leq \lim_{\xi \to \infty} \sum_{\zeta=1}^{\xi} \sup_{\boldsymbol{S}} \left[\prod_{r=1}^{\zeta} \mathbb{P}_1^{\boldsymbol{S}} \left(\frac{\sum_{j=(r-1)n_b+1}^{rn_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon \right) \right] \\
&\leq \lim_{\xi \to \infty} \sum_{\zeta=1}^{\xi} \prod_{r=1}^{\zeta} \left[\sup_{\boldsymbol{S}} \mathbb{P}_1^{\boldsymbol{S}} \left(\frac{\sum_{j=(r-1)n_b+1}^{rn_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon \right) \right] \\
&= \lim_{\xi \to \infty} \sum_{\zeta=1}^{\xi} \left[\sup_{\boldsymbol{S}} \mathbb{P}_1^{\boldsymbol{S}} \left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon \right) \right]^{\zeta}. \qquad \text{(B.77)}
\end{aligned}
$$

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

$$
I_{\boldsymbol{S},b} \triangleq \mathbb{E}_1^{\boldsymbol{S}} \left[\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} \right] = \frac{\sum_{j=1}^{n_b} \mathbb{E}_{p_{\boldsymbol{S}[j]}} [Z[j]]}{n_b} \geq \overset{*}{I}, \qquad \text{(B.78)}
$$

where the inequality follows from Lemma 6. This in turn implies that for any $\boldsymbol{S}$ we have that

$$\mathbb{P}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon\right) = \mathbb{P}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} < \overset{*}{I} - \epsilon + I_{\boldsymbol{S},b} - I_{\boldsymbol{S},b}\right)$$
$$\leq \mathbb{P}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} < I_{\boldsymbol{S},b} - \epsilon\right)$$
$$\leq \mathbb{P}_1^{\boldsymbol{S}}\left(\left|\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} - I_{\boldsymbol{S},b}\right| > \epsilon\right). \tag{B.79}$$

Define

$$\bar{\sigma}^2 \triangleq \max_{\boldsymbol{E} \in \mathcal{E}} \mathrm{Var}_{p_{\boldsymbol{E}}}\left(Z[1]\right). \tag{B.80}$$

From eq. (3.25), we have that $\bar{\sigma}^2 < \infty$. Then, by Chebychev's inequality

$$\mathbb{P}_1^{\boldsymbol{S}}\left(\left|\frac{\sum_{j=1}^{n_b} Z[j]}{n_b} - I_{\boldsymbol{S},b}\right| > \epsilon\right) \leq \mathrm{Var}_1^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{n_b} Z[j]}{n_b}\right)\frac{1}{\epsilon^2}$$
$$= \frac{1}{\epsilon^2 n_b^2}\sum_{j=1}^{n_b} \mathrm{Var}_{p_{\boldsymbol{S}[j]}}\left(Z[j]\right)$$
$$\leq \frac{\sum_{j=1}^{n_b} \bar{\sigma}^2}{n_b^2 \epsilon^2}$$
$$= \frac{\bar{\sigma}^2}{n_b \epsilon^2}. \tag{B.81}$$

By using (B.74), (B.77), (B.79) and (B.81) we then have that

$$\sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}}\left[\frac{\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b}\right] \leq 1 + \lim_{\xi\to\infty}\sum_{\zeta=1}^{\xi}\left[\frac{\bar{\sigma}^2}{n_b\epsilon^2}\right]^{\zeta}. \tag{B.82}$$

Let $0 < \delta < 1$. Since n_b is increasing with b, we have that for all $b > \tilde{b}$, where

$\tilde{b}$ large enough

$$\sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}} \left[\frac{\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b} \right] \leq 1 + \lim_{\xi \to \infty} \sum_{\zeta=1}^{\xi} \delta^{\zeta} = \sum_{\zeta=0}^{\infty} \delta^{\zeta} = \frac{1}{1-\delta} \tag{B.83}$$

which implies that for all $b > \tilde{b}$

$$\sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}} \left[\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) \right] \leq \frac{b}{(\overset{*}{I} - \epsilon)(1-\delta)}. \tag{B.84}$$

Since (B.84) holds for all ϵ small enough we have that

$$\sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}} \left[\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) \right] \leq \frac{b}{\overset{*}{I}(1-\delta)}. \tag{B.85}$$

Finally, since $\delta \to 0$ as $b \to \infty$ we have that

$$\text{WADD}(\tau_M(\overset{*}{\boldsymbol{\alpha}}, b)) = \sup_{\boldsymbol{S}} \mathbb{E}_1^{\boldsymbol{S}} \left[\tau_M(\overset{*}{\boldsymbol{\alpha}}, b) \right] \leq \frac{b}{\overset{*}{I}}(1 + o(1)) \tag{B.86}$$

as $b \to \infty$.

APPENDIX C

PROOFS FOR CHAPTER 4

C.1 Useful Notation

For the theoretical analysis of Chapter 4, we focus on the case of two post-change phases (one transient phase and one persistent phase). The results in this chapter hold for the case of arbitrary number of phases $K \geq 2$ known by the decision maker, but in that case the analysis becomes cumbersome.

Consider the sequences $\boldsymbol{S}^{(1)} \triangleq \{\boldsymbol{S}^{(1)}[k]\}_{k=1}^{\infty}$ and $\boldsymbol{S}^{(2)} \triangleq \{\boldsymbol{S}^{(2)}[k]\}_{k=1}^{\infty}$ which characterize the location of the anomalous nodes at each time instant for post-change phase 1 and 2 respectively. For clarity of notation we will use ν in the Appendix C to denote the first changepoint ν_1, d to denote the transient duration d_1 and ρ to denote ρ_1. Then, for fixed changepoint ν and transient duration d (second changepoint $\nu_2 = \nu + d$), $\nu \geq 1$, $d \geq 0$, we have the following statistical model for the two-phase case:

$$\boldsymbol{X}[k] \sim \begin{cases} g\left(\boldsymbol{X}[k]\right), & 1 \leq k < \nu, \\ p_{\boldsymbol{S}^{(1)}[k]}\left(\boldsymbol{X}[k]\right), & \nu \leq k < \nu + d, \\ p_{\boldsymbol{S}^{(2)}[k]}\left(\boldsymbol{X}[k]\right), & k \geq \nu + d. \end{cases} \tag{C.1}$$

Furthermore, 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

$$\Gamma_{\boldsymbol{S}}(k, \nu_1, \nu_2) \triangleq \begin{cases} \left[\prod_{j=\nu_1}^{\min\{\nu_2-1,k\}} \prod_{\ell \in \boldsymbol{S}_1[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])} \right] \cdot \left[\prod_{j=\nu_2}^{k} \prod_{\ell \in \boldsymbol{S}_2[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])} \right], & \nu_1 < \nu_2, \\ \prod_{j=\nu_1}^{k} \prod_{\ell \in \boldsymbol{S}_2[j]} \frac{f_\ell(X_\ell[j])}{g_\ell(X_\ell[j])}, & \nu_1 = \nu_2. \end{cases} \tag{C.2}$$

In addition, for the model in (4.7), 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

$$\mathcal{L}_{\boldsymbol{\alpha}}(k,\nu_1,\nu_2) = \begin{cases} \left[\prod_{j=\nu_1}^{\min\{\nu_2-1,k\}} \frac{\overline{p}^{(1)}_{\boldsymbol{\alpha}^{(1)}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}\right] \cdot \left[\prod_{j=\nu_2}^{k} \frac{\overline{p}^{(2)}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}\right], & \nu_1 \le \nu_2, \\ \prod_{j=\nu_1}^{k} \frac{\overline{p}^{(2)}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{X}[j])}{g(\boldsymbol{X}[j])}, & \nu_1 = \nu_2, \end{cases} \tag{C.3}$$

and the log-likelihood ratio at phase $i \in [2]$, time k for $\boldsymbol{\alpha} = \mathring{\boldsymbol{\alpha}}$ by

$$Z^{(i)}[k] = \log \frac{\overline{p}^{(i)}_{\mathring{\boldsymbol{\alpha}}^{(i)}}(\boldsymbol{X}[k])}{g(\boldsymbol{X}[k])}. \tag{C.4}$$

Finally, denote the logarithm of the weighted likelihood ratio [35] in (C.3) for $\rho \in (0,1)$ by

$$\omega_{\boldsymbol{\alpha}}(k,\nu_1,\nu_2) = \begin{cases} \log \frac{\left(\prod_{j=\nu_1}^{\min\{\nu_2-1,k\}} \overline{p}^{(1)}_{\boldsymbol{\alpha}^{(1)}}(\boldsymbol{X}[j])(1-\rho)\right) \rho^{\mathbb{1}\{k\ge v_2\}} \prod_{j=\nu_2}^{k} \overline{p}^{(2)}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{X}[j])}{\prod_{j=\nu_1}^{k} g(\boldsymbol{X}[j])}, & \nu_1 \le \nu_2, \\ \log \frac{\rho \prod_{j=\nu_1}^{k} \overline{p}^{(2)}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{X}[j])}{\prod_{j=\nu_1}^{k} g(\boldsymbol{X}[j])}, & \nu_1 = \nu_2. \end{cases} \tag{C.5}$$

C.2 Useful Lemma

The proofs of Chapter 4 rely on the following lemma:

Lemma 10. *For any stopping rule τ, define its truncated version by $\tau^{(N)} \triangleq \min\{\tau, N\}$ where N is a positive integer. Then, we have that for any $d \ge 0$*

$$\mathrm{WADD}_d(\tau) \ge \mathrm{WADD}_d(\tau^{(N)}). \tag{C.6}$$

Proof. Fix $N \ge 1$. Consider initially the case that $N \ge \nu$. Then, since $\{\tau^{(N)} \ge \nu\} = \{\min\{\tau, N\} \ge \nu\} = \{\tau \ge \nu\} \cap \{N \ge \nu\}$, we have that $\{\tau^{(N)} \ge \nu\} = \{\tau \ge \nu\}$. Since $\tau^{(N)} \le \tau$, this implies that for any $N \ge \nu$ and any $\boldsymbol{S}$, d we have that

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)} - \nu + 1|\tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]] \\
&= \mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)} - \nu + 1|\tau \geq \nu, \boldsymbol{X}[1, \nu - 1]] \\
&\leq \mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau - \nu + 1|\tau \geq \nu, \boldsymbol{X}[1, \nu - 1]]. \qquad \text{(C.7)}
\end{aligned}$$

For the case that $N < \nu$, we have that that $\mathbb{P}^{\boldsymbol{S}}_{\nu,d}(\tau^{(N)} \geq \nu) = 0$, which implies that for any $N < \nu$ and any $\boldsymbol{S}$, d we have that

$$\mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\tau^{(N)} - \nu + 1|\tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] = 1, \qquad \text{(C.8)}$$

by convention. Furthermore, note that for any $\boldsymbol{S}$, d we have that

$$\mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\tau - \nu + 1|\tau \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \geq 1. \qquad \text{(C.9)}$$

From (C.7) - (C.9), we have that for any $\nu \geq 1$ and any $\boldsymbol{S}$, d

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)} - \nu + 1|\tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]] \\
&\leq \mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau - \nu + 1|\tau \geq \nu, \boldsymbol{X}[1, \nu - 1]]. \qquad \text{(C.10)}
\end{aligned}$$

By taking the sup and ess sup on both sides the lemma is established. □

Remark 8. *It should be noted that the use of the superscript in* $\tau^{(N)}$ *of Lemma 10 (as well as Lemma 8) is not related to the superscript used to denote post-change phases, which appears in Chapter 4, as well as Appendix C. In particular, with some abuse of notation, it is the convention in this dissertation that when a superscript in the form of* $(\cdot)$ *is used on a stopping time it refers to a truncated stopping time. On the contrary, when it is used in any other quantity aside from a stopping time it is used to denote which post-change phase said quantity is related to.*

C.3 Proof of Lemma 7

Let $N \geq 1$. For any stopping rule τ, we have that from Lemma 10 for any ν, d, $N \geq 1$

$$\mathrm{WADD}_d(\tau) \geq \mathrm{WADD}_d(\tau^{(N)}) \geq \sup_{\boldsymbol{S}} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]]. \tag{C.11}$$

Following, we have that for any $\nu, d, \boldsymbol{S}$ and $N > \nu + d$

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{S}}_{\nu,d}[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]] \\
&= \mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\sum_{j=\nu}^{\infty} \mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\overset{(a)}{=} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\sum_{j=\nu}^{N} \mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&= \sum_{j=\nu}^{N} \mathbb{E}^{\boldsymbol{S}}_{\nu,d}\left[\mathbb{1}_{\{\tau^{(N)} \geq j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\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| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\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| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&= \mathbb{E}_{\infty}\left[\Gamma_{\boldsymbol{S}}\left(\nu-1, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} \geq \nu\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\quad + \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| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&= \mathbb{E}_{\infty}\left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\quad + \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| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\overset{(c)}{=} \mathbb{E}_{\infty}\left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right] \\
&\quad + \mathbb{E}_{\infty}\left[\sum_{j=\nu}^{N-1} \Gamma_{\boldsymbol{S}}\left(j, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu - 1]\right],
\end{aligned} \tag{C.12}$$

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

$$\begin{aligned}
&\sup_{\boldsymbol{S}} \mathbb{E}_{\nu,d}^{\boldsymbol{S}}[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1]] \\
&\overset{(d)}{=} \mathbb{E}_\infty \left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&+ \sup_{\boldsymbol{S}} \mathbb{E}_\infty \left[\sum_{j=\nu}^{N-1} \Gamma_{\boldsymbol{S}}\left(j, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&\overset{(e)}{=} \mathbb{E}_\infty \left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&+ \sup_{\boldsymbol{S}^{(1)}[1,N-1], \boldsymbol{S}^{(2)}[1,N-1]} \mathbb{E}_\infty \left[\sum_{j=\nu}^{N-1} \Gamma_{\boldsymbol{S}}\left(j, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&\overset{(f)}{=} \mathbb{E}_\infty \left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&+ \sup_{\boldsymbol{S}^{(1)}[\nu,\nu+d-1], \boldsymbol{S}^{(2)}[\nu+d,N-1]} \mathbb{E}_\infty \left[\sum_{j=\nu}^{N-1} \Gamma_{\boldsymbol{S}}\left(j, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right]
\end{aligned} \tag{C.13}$$

where (d) follows because the first term in (C.12) does not depend on $\boldsymbol{S}$; (e) follows since the summation in the second expectation is from $j = \nu$ to $N-1$ which implies that only the first $N-1$ samples are involved in the calculation of $\Gamma_{\boldsymbol{S}}\left(j, \nu, \nu+d\right)$; and (f) follows from the definitions of changepoints ν and $\nu + d$. By following similar steps to Appendix B.2, i.e., using the fact that the sup can be lower bounded by the average, it can be shown that for any $\boldsymbol{\alpha}$

$$\begin{aligned}
&\sup_{\boldsymbol{S}^{(1)}[\nu,\nu+d-1], \boldsymbol{S}^{(2)}[\nu+d,N-1]} \mathbb{E}_\infty \left[\sum_{j=\nu}^{N-1} \Gamma_{\boldsymbol{S}}\left(j, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right] \\
&\geq \mathbb{E}_\infty \left[\sum_{j=\nu}^{N-1} \mathcal{L}_{\boldsymbol{\alpha}}\left(j, \nu, \nu+d\right) \mathbb{1}_{\{\tau^{(N)} > j\}} \middle| \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1] \right].
\end{aligned} \tag{C.14}$$

Then, by combining (C.11), (C.13) and (C.14) we have that for any b, $\boldsymbol{\alpha}$

$$
\begin{aligned}
&\mathrm{WADD}_d(\tau) \\
&\geq \mathbb{E}_\infty\left[\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&\quad+\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|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&=\mathbb{E}_\infty\left[\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&\quad+\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\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&=\mathbb{E}_\infty\left[\mathbb{1}_{\{\tau^{(N)}\geq\nu\}}\mathcal{L}_{\boldsymbol{\alpha}}(\nu-1,\nu,\nu+d)\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&\quad+\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\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\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\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\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\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&=\sum_{j=\nu}^{N}\overline{\mathbb{E}}_{\nu,d}^{\boldsymbol{\alpha}}\left[\mathbb{1}_{\{\tau^{(N)}\geq j\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&=\overline{\mathbb{E}}_{\nu,d}^{\boldsymbol{\alpha}}\left[\sum_{j=\nu}^{N}\mathbb{1}_{\{\tau^{(N)}\geq j\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&=\overline{\mathbb{E}}_{\nu,d}^{\boldsymbol{\alpha}}\left[\sum_{j=\nu}^{\infty}\mathbb{1}_{\{\tau^{(N)}\geq j\}}\middle|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]\right] \\
&=\overline{\mathbb{E}}_{\nu,d}^{\boldsymbol{\alpha}}[\tau^{(N)}-\nu+1|\tau^{(N)}\geq\nu,\boldsymbol{X}[1,\nu-1]]. \qquad \text{(C.15)}
\end{aligned}
$$

From the Monotone Convergence Theorem, since $\tau^{(N)} - \nu + 1$ and $\mathbb{1}_{\{\tau^{(N)} \geq \nu\}}$ are non-decreasing with N, we have that

$$\begin{aligned}
&\lim_{N\to\infty} \overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}[\tau^{(N)} - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1]] \\
&= \lim_{N\to\infty} \frac{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[(\tau^{(N)} - \nu + 1)\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} | \boldsymbol{X}[1, \nu-1]\right]}{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} | \boldsymbol{X}[1, \nu-1]\right]} \\
&= \frac{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\lim_{N\to\infty}(\tau^{(N)} - \nu + 1)\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} | \boldsymbol{X}[1, \nu-1]\right]}{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\lim_{N\to\infty}\mathbb{1}_{\{\tau^{(N)} \geq \nu\}} | \boldsymbol{X}[1, \nu-1]\right]} \\
&= \frac{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[(\tau - \nu + 1)\mathbb{1}_{\{\tau \geq \nu\}} | \boldsymbol{X}[1, \nu-1]\right]}{\overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\mathbb{1}_{\{\tau \geq \nu\}} | \boldsymbol{X}[1, \nu-1]\right]} \\
&= \overline{\mathbb{E}}^{\boldsymbol{\alpha}}_{\nu,d}\left[\tau - \nu + 1 | \tau^{(N)} \geq \nu, \boldsymbol{X}[1, \nu-1]\right]. \qquad \text{(C.16)}
\end{aligned}$$

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

$$\mathrm{WADD}_d(\tau) \geq \overline{\mathrm{WADD}}_{\boldsymbol{\alpha},d}(\tau). \qquad \text{(C.17)}$$

C.4 Proof of Theorem 11

Our upper bound analysis is based on the proof technique in [35]. In particular, it can be shown that the M-WD-CUSUM test statistic is given by

$$\Omega_{\boldsymbol{\alpha}}[k] = \max_{1 \leq i_1 \leq i_2 \leq k+1} \omega_{\boldsymbol{\alpha}}(k, i_1, i_2). \qquad \text{(C.18)}$$

In addition, due to the Markov property and recursive structure of the M-WD-CUSUM test statistic, we have that for any $\boldsymbol{S}$, $\boldsymbol{\alpha}$ and any values of b and d

$$\mathrm{WADD}_d(\tau_\Omega(\boldsymbol{\alpha}, b)) = \sup_{\boldsymbol{S}} \mathbb{E}^{\boldsymbol{S}}_{1,d}\left[\tau_\Omega(\boldsymbol{\alpha}, b)\right]. \qquad \text{(C.19)}$$

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

$$d \sim c_1' \frac{b}{\overset{*}{I}^{(1)} + \log(1-\rho)} \tag{C.20}$$

as $b \to \infty$. Depending on the value of c_1' we can proceed to bound $\sup_{\boldsymbol{S}} \mathbb{E}_{1,\boldsymbol{d}}^{\boldsymbol{S}} [\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)]$ as in [35].

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

$$n_b \triangleq \frac{b}{\overset{*}{I}^{(1)} + \log(1-\rho) - \epsilon}, \tag{C.21}$$

$$c_\epsilon \triangleq \left\lfloor c_1' \frac{1-\epsilon}{1+\epsilon} \right\rfloor. \tag{C.22}$$

We then have that

$$\begin{aligned}
&\sup_{\boldsymbol{S}} \mathbb{E}_{1,d}^{\boldsymbol{S}} \left[\frac{\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b} \right] \\
&\overset{\text{(a)}}{=} \sup_{\boldsymbol{S}} \int_0^\infty \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(\frac{\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b} > x \right) dx \\
&\overset{\text{(b)}}{\le} \sup_{\boldsymbol{S}} \sum_{\zeta=0}^{\infty} \mathbb{P}_{1,d}^{\boldsymbol{S}} (\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&\le 1 + \sum_{\zeta=1}^{c_\epsilon} \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} (\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) + \sup_{\boldsymbol{S}} \sum_{\zeta=c_\epsilon+1}^{\infty} \mathbb{P}_{1,d}^{\boldsymbol{S}} (\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&\le 1 + \sum_{\zeta=1}^{c_\epsilon} \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} (\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) + \lim_{\xi \to \infty} \sum_{\zeta=c_\epsilon+1}^{\xi} \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} (\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b),
\end{aligned} \tag{C.23}$$

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.

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 the first post-change phase. As a result, we have that for any $\boldsymbol{S}$ and $\zeta \in [c_\epsilon]$

$$
\begin{aligned}
&\mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \\
&= \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\max_{1 \le k \le \zeta n_b} \Omega_{\overset{*}{\boldsymbol{\alpha}}}[k] < b\right) \\
&\overset{(c)}{=} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\max_{1 \le k \le \zeta n_b} \max_{1 \le i_1 \le i_2 \le k+1} \omega_{\overset{*}{\boldsymbol{\alpha}}}(k, i_1, i_2) < b\right) \\
&\overset{(d)}{\le} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\omega_{\overset{*}{\boldsymbol{\alpha}}}(r n_b, (r-1) n_b + 1, d+1) < b, \forall\, r \in [\zeta]\right) \\
&\overset{(e)}{=} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\sum_{j=(r-1)n_b+1}^{r n_b} \big(Z^{(1)}[j] + \log(1-\rho)\big) < b,\ \forall\, r \in [\zeta]\right) \\
&\overset{(f)}{=} \prod_{r=1}^{\zeta} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{1}{n_b} \sum_{j=(r-1)n_b+1}^{r n_b} \big(Z^{(1)}[j] + \log(1-\rho)\big) < \frac{b}{n_b}\right) \\
&\overset{(g)}{=} \prod_{r=1}^{\zeta} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=(r-1)n_b+1}^{r n_b} Z^{(1)}[j]}{n_b} < \overset{*}{I}^{(1)} - \epsilon\right)
\end{aligned} \tag{C.24}
$$

where (c) follows from (C.18); (d) follows by binning the observations and bounding the maxima (see [22] and [35]); (e) follows from (C.5); (f) follows from independence of data across times *conditioned* on $\boldsymbol{S}$; and (g) follows from the definition of n_b. We then have that for any $b > 0$ from (C.23) and (C.24)

$$
\begin{aligned}
\sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) &\le \sup_{\boldsymbol{S}} \left[\prod_{r=1}^{\zeta} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=(r-1)n_b+1}^{r n_b} Z^{(1)}[j]}{n_b} < \overset{*}{I}^{(1)} - \epsilon\right)\right] \\
&= \left[\sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b} < \overset{*}{I}^{(1)} - \epsilon\right)\right]^{\zeta}.
\end{aligned} \tag{C.25}
$$

By Lemma 6 of Chapter 3 we have that

$$I_{\boldsymbol{S},b,d} \triangleq \mathbb{E}^{\boldsymbol{S}}_{1,d}\left[\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b}\right] = \frac{\sum_{j=1}^{n_b} \mathbb{E}_{p_{\boldsymbol{S}^{(1)}[j]}}\left[Z^{(1)}[j]\right]}{n_b} \geq \overset{*}{I}{}^{(1)}. \tag{C.26}$$

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

$$\begin{aligned}
\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b} < \overset{*}{I}{}^{(1)} - \epsilon\right) &= \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b} < \overset{*}{I}{}^{(1)} - \epsilon + I_{\boldsymbol{S},b,d} - I_{\boldsymbol{S},b,d}\right) \\
&\leq \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b} < I_{\boldsymbol{S},b,d} - \epsilon\right) \\
&\leq \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\left|\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b} - I_{\boldsymbol{S},b,d}\right| > \epsilon\right). \qquad \text{(C.27)}
\end{aligned}$$

Define

$$(\bar{\sigma}^{(1)})^2 \triangleq \max_{\boldsymbol{E} \in \mathcal{E}^{(1)}} \mathrm{Var}_{p_{\boldsymbol{E}}}\left(Z^{(1)}[1]\right). \tag{C.28}$$

From eq. (4.20), we have that $(\bar{\sigma}^{(1)})^2 < \infty$. Then, by Chebychev's inequality

$$\begin{aligned}
\mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\left|\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b} - I_{\boldsymbol{S},b,d}\right| > \epsilon\right) &\leq \mathrm{Var}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{n_b} Z^{(1)}[j]}{n_b}\right)\frac{1}{\epsilon^2} \\
&= \frac{1}{\epsilon^2 n_b^2}\sum_{j=1}^{n_b} \mathrm{Var}_{p_{\boldsymbol{S}^{(1)}[j]}}\left(Z^{(1)}[j]\right) \\
&\leq \frac{\sum_{j=1}^{n_b} (\bar{\sigma}^{(1)})^2}{n_b^2 \epsilon^2} \\
&= \frac{(\bar{\sigma}^{(1)})^2}{n_b \epsilon^2} \\
&\leq \delta \qquad \text{(C.29)}
\end{aligned}$$

for large b, which from (C.25) implies that for large b

$$\sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \leq \delta^\zeta. \tag{C.30}$$

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 the transient or the persistent anomaly size phase. Define

$$t \triangleq \left\lceil \frac{\overset{*}{I}^{(1)}}{\min\{\overset{*}{I}^{(1)}, \overset{*}{I}^{(2)}\}} \right\rceil + 1. \tag{C.31}$$

We then have that for large b, $c_\epsilon n_b \leq \nu_2 \leq (c_\epsilon + t)n_b$. Consider ζ such that $c_\epsilon + (l-1)t \leq \zeta \leq c_\epsilon + lt - 1$, for any $l \geq 1$. We then have that

$$\begin{aligned}
\sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) &= \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(\max_{1 \leq k \leq \zeta n_b} \Omega_{\overset{*}{\boldsymbol{\alpha}}}[k] < b \right) \\
&\leq \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(A_1 \cap A_2 \right) \\
&\leq \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(A_1 \right) \cdot \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(A_2 \right)
\end{aligned} \tag{C.32}$$

where

$$A_1 \triangleq \{ w_{\overset{*}{\boldsymbol{\alpha}}} (rn_b, (r-1)n_b + 1, d+1) < b, \forall\, r \in [c_\epsilon] \} \tag{C.33}$$

$$A_2 \triangleq \{ w_{\overset{*}{\boldsymbol{\alpha}}} ((c_\epsilon + rt)n_b, (c_\epsilon + (r-1)t)n_b + 1, d+1) < b, \forall\, r \in [l-1] \}. \tag{C.34}$$

By following similar steps to (C.24) - (C.30) it can be established that for large b

$$\sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(A_1 \right) \leq \delta^{c_\epsilon} \tag{C.35}$$

and

$$\sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}} \left(A_2 \right) \leq \delta^{l-1}. \tag{C.36}$$

We then have than from (C.23), (C.30), (C.32), (C.35), (C.36) and the definition of t

$$\begin{aligned}\sup_{\boldsymbol{S}} \mathbb{E}_{1,d}^{\boldsymbol{S}}\left[\frac{\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)}{n_b}\right] &\le 1+\sum_{\zeta=1}^{c_\epsilon}\delta^\zeta+\lim_{\xi\to\infty}\sum_{l=1}^{\xi} t\delta^{c_\epsilon+l-1}\\ &= \frac{1}{1-\delta}+t\delta^{c_\epsilon}+(t-1)\delta^{c_\epsilon+1}\frac{1}{1-\delta}\\ &\triangleq 1+\delta' \end{aligned} \tag{C.37}$$

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

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)) = \sup_{\boldsymbol{S}} \mathbb{E}_{1,d}^{\boldsymbol{S}}\left[\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)\right] \le \frac{b}{\mathring{I}^{(1)}}(1+o(1)). \tag{C.38}$$

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

$$\begin{aligned} n_b' &= \left(d+\frac{b-\log\rho-d(\mathring{I}^{(1)}+\log(1-\rho))}{\mathring{I}^{(2)}}\right)(1+\epsilon)\\ &\sim b\left(\frac{c_1'}{\mathring{I}^{(1)}}+\frac{1-c_1'}{\mathring{I}^{(2)}}\right)(1+\epsilon). \end{aligned}\tag{C.39}$$

This implies that

$$\lim_{b\to\infty}\frac{n_b'}{d} = \left(1+\left(\frac{1}{c_1'}-1\right)\frac{\mathring{I}^{(1)}}{\mathring{I}^{(2)}}\right)(1+\epsilon) > 1 \tag{C.40}$$

which in turn implies that for large b, $n_b' > d$ and $n_b' - d \to \infty$ as $b\to\infty$ [35].

By analyzing the expectation as in case 1 we have that

$$\begin{aligned} &\sup_{\boldsymbol{S}} \mathbb{E}_{1,d}^{\boldsymbol{S}}\left[\frac{\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)}{n_b'}\right]\\ &\le 1+\sup_{\boldsymbol{S}}\mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)>n_b')+\sup_{\boldsymbol{S}}\sum_{\zeta=2}^{\infty}\mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)>\zeta n_b')\\ &\le 1+\sup_{\boldsymbol{S}}\mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)>n_b')+\lim_{\xi\to\infty}\sum_{\zeta=2}^{\xi}\sup_{\boldsymbol{S}}\mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\mathring{\boldsymbol{\alpha}},b)>\zeta n_b'). \end{aligned}\tag{C.41}$$

For fixed $\boldsymbol{S}$ and since for any constants x, y and random variables X, Y,

$\mathbb{P}(X+Y<x+y) \leq \mathbb{P}(X<x)+\mathbb{P}(Y<y)$ we then have that

$$\begin{aligned}
&\mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > n_b') \\
&= \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\max_{1\leq k\leq n_b'} \Omega_{\overset{*}{\boldsymbol{\alpha}}}[k] < b\right) \\
&\leq \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(w_{\overset{*}{\boldsymbol{\alpha}}}[n_b', 1, d+1] < b\right) \\
&= \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\sum_{j=1}^{d}\left(Z^{(1)}[j]+\log(1-\rho)\right) + \log\rho + \sum_{j=d+1}^{n_b'} Z^{(2)}[j] < b\right) \\
&= \mathbb{P}_{1,d}^{\boldsymbol{S}}\Bigg(\sum_{j=1}^{d}\left(Z^{(1)}[j]+\log(1-\rho)\right) + \sum_{j=d+1}^{n_b'} Z^{(2)}[j] \\
&\qquad\qquad < d(\overset{*}{I}^{(1)}+\log(1-\rho)) + (n_b'-d)\overset{*}{I}^{(2)} - \epsilon a\Bigg) \\
&< \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\sum_{j=1}^{d}\left(Z^{(1)}[j]+\log(1-\rho)\right) < d(\overset{*}{I}^{(1)}+\log(1-\rho)) - \frac{\epsilon a}{2}\right) \\
&\quad + \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\sum_{j=d+1}^{n_b'} Z^{(2)}[j] < (n_b'-d)\overset{*}{I}^{(2)} - \frac{\epsilon a}{2}\right) \\
&\leq \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} < \overset{*}{I}^{(1)} - \frac{\epsilon a}{2d}\right) + \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\frac{\sum_{j=d+1}^{n_b'} Z^{(2)}[j]}{n_b'-d} < \overset{*}{I}^{(2)} - \frac{\epsilon a}{2(n_b'-d)}\right)
\end{aligned} \tag{C.42}$$

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

$$\begin{aligned}
\sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b) \leq &\sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} < \overset{*}{I}^{(1)} - \frac{\epsilon a}{2d}\right) \\
&+ \sup_{\boldsymbol{S}} \mathbb{P}_{1,d}^{\boldsymbol{S}}\left(\frac{\sum_{j=d+1}^{n_b'} Z^{(2)}[j]}{n_b'-d} < \overset{*}{I}^{(2)} - \frac{\epsilon a}{2(n_b'-d)}\right).
\end{aligned} \tag{C.43}$$

Following, we upper bound both of the terms in the right hand side of (C.43). In particular, from (C.26) we have that

$$\begin{aligned}
&\sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} < \mathring{I}^{(1)} - \frac{\epsilon a}{2d}\right) \\
&= \sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} < \mathring{I}^{(1)} - \frac{\epsilon a}{2d} + I_{\boldsymbol{S},b,d} - I_{\boldsymbol{S},b,d}\right) \\
&\leq \sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} < I_{\boldsymbol{S},b,d} - \frac{\epsilon a}{2d}\right) \\
&\leq \sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\left|\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} - I_{\boldsymbol{S},b,d}\right| > \frac{\epsilon a}{2d}\right). \qquad \text{(C.44)}
\end{aligned}$$

From Chebychev's inequality we then have that

$$\begin{aligned}
\sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\left|\frac{\sum_{j=1}^{d} Z^{(1)}[j]}{d} - I_{\boldsymbol{S},b,d}\right| > \frac{\epsilon a}{2d}\right) &\leq \sup_{\boldsymbol{S}} \frac{1}{d^2}\mathrm{Var}^{\boldsymbol{S}}_{1,d}\left(\sum_{j=1}^{d} Z^{(1)}[j]\right)\left(\frac{2d}{\epsilon a}\right)^2 \\
&\leq \frac{1}{d}\left(\frac{2d\bar{\sigma}^{(1)}}{\epsilon a}\right)^2 \\
&\leq \frac{\delta}{2} \qquad \text{(C.45)}
\end{aligned}$$

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

$$\sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}\left(\frac{\sum_{j=d+1}^{n'_b} Z^{(2)}[j]}{n'_b - d} < \mathring{I}^{(2)} - \frac{\epsilon a}{2(n'_b - d)}\right) \leq \frac{\delta}{2} \qquad \text{(C.46)}$$

for large b.

Define

$$t' \triangleq \left\lceil \frac{1}{\left(\frac{c_1'}{\overset{*}{I}^{(1)}} + \frac{1-c_1'}{\overset{*}{I}^{(2)}}\right)\min\{\overset{*}{I}^{(1)}, \overset{*}{I}^{(2)}\}} \right\rceil + 1. \tag{C.47}$$

Following similar arguments to (C.32) - (C.36) we can establish that if $(l-1)t+1 \leq \zeta \leq lt$ for any $l \geq 1$ we then we have that

$$\sup_{\boldsymbol{S}} \mathbb{P}^{\boldsymbol{S}}_{1,d}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b) > \zeta n_b') \leq t'\delta^l. \tag{C.48}$$

Combining (C.41), (C.45), (C.46) and (C.48) we have that

$$\begin{aligned}
\sup_{\boldsymbol{S}} \mathbb{E}^{\boldsymbol{S}}_{1,d}\left[\frac{\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)}{n_b}\right] &\leq 1 + \delta + \lim_{\xi\to\infty}\sum_{\zeta=2}^{\xi} t'\delta^{\zeta-1} \\
&= \frac{1}{1-\delta} + t'\delta + (t'-1)\frac{\delta^2}{1-\delta} \\
&\triangleq \delta''
\end{aligned} \tag{C.49}$$

where $\delta'' \to 0$ as $b \to \infty$. As a result, we have that from (C.39) and (C.49)

$$\mathrm{WADD}_{\boldsymbol{d}}(\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)) = \sup_{\boldsymbol{S}} \mathbb{E}^{\boldsymbol{S}}_{1,d}\left[\tau_\Omega(\overset{*}{\boldsymbol{\alpha}}, b)\right] \leq b\left(\frac{c_1'}{\overset{*}{I}^{(1)}} + \frac{1-c_1'}{\overset{*}{I}^{(2)}}\right)(1+o(1)). \tag{C.50}$$

Finally, from (C.38) and (C.50) the theorem is established.

REFERENCES

[1] 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. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 2946–2950.

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

[3] G. Rovatsos, S. Zou, A. D. Domínguez-García, and V. V. Veeravalli, "Stochastic control of power supply in data centers," in *Proc. 52nd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Nov. 2018, pp. 1470–1474.

[4] K. Mechitov, W. Kim, G. Agha, and T. Nagayama, "High-frequency distributed sensing for structure monitoring," in *Proc. First Int. Workshop on Networked Sensing Systems (INSS)*, Tokyo, Japan, Jun. 2004, pp. 2787–2797.

[5] J. Rice, K. Mechitov, S. Sim, T. Nagayama, S. Jang, R. Kim, B. Spencer, G. Agha, and Y. Fujino, "Flexible smart sensor framework for autonomous structural health monitoring," *Smart Structures and Systems*, vol. 6, no. 5-6, pp. 423–438, Jul. 2010.

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

[7] S. E. Fienberg and G. Shmueli, "Statistical issues and challenges associated with rapid detection of bio-terrorist attacks," *Statistics in Medicine*, vol. 24, no. 4, pp. 513–529, Feb. 2005.

[8] M. Frisn, "Optimal sequential surveillance for finance, public health, and other areas," *Sequential Analysis*, vol. 28, no. 3, pp. 310–337, Jul. 2009.

[9] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. Glob. Tel. Conf. (GLOBECOM)*, New Orleans, USA, Nov. 2008, pp. 1–5.

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

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

[12] H. V. Poor and O. Hadjiliadis, *Quickest Detection.* Cambridge, UK: Cambridge University Press, Dec. 2009.

[13] W. A. Shewhart, "The application of statistics as an aid in maintaining quality of a manufactured product," *Journal of the American Statistical Association*, vol. 20, no. 152, pp. 546–548, 1925.

[14] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, Jun. 1954.

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

[16] A. N. Shiryaev, *Optimal Stopping Rules.* Springer-Verlag, 1978.

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

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

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

[20] G. V. Moustakides, "Quickest detection of abrupt changes for a class of random processes," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1965–1968, Sep. 1998.

[21] H. V. Poor, "Quickest detection with exponential penalty for delay," *Annals of Statistics*, vol. 26, no. 6, pp. 2179–2205, Dec. 1998.

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

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

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

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

[26] 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.

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

[28] 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.

[29] 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.

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

[31] 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.

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

[33] S. Zou, V. V. Veeravalli, J. Li, D. Towsley, and A. Swami, "Distributed quickest detection of significant events in networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 8454–8458.

[34] J. Li, D. Towsley, S. Zou, V. V. Veeravalli, and G. Ciocarlie, "A consensus-based approach for distributed quickest detection of significant events in networks," in *Proc. IEEE 53rd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Nov 2019, pp. 1–4.

[35] 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.

[36] C.-D. Fuh, "SPRT and CUSUM in hidden Markov models," *Annals of Statistics*, vol. 31, no. 3, pp. 942–977, Jun. 2003.

[37] C.-D. Fuh, "Asymptotic operating characteristics of an optimal change point detection in hidden Markov models," *Annals of Statistics*, vol. 32, no. 5, pp. 2305–2339, Mar. 2004.

[38] C.-D. Fuh and Y. Mei, "Quickest change detection and Kullback-Leibler divergence for two-state hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 63, pp. 4866–4878, Sep. 2015.

[39] C.-D. Fuh and A. G. Tartakovsky, "Asymptotic Bayesian theory of quickest change detection for hidden Markov models," *IEEE Trans. Inform. Theory*, vol. 65, no. 1, pp. 511–529, Jan. 2019.

[40] G. V. Moustakides, "Detecting changes in hidden Markov models," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 2394–2398.

[41] 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, USA, Mar. 2017.

[42] G. V. Moustakides and V. V. Veeravalli, "Sequentially detecting transitory changes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

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

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

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

[46] S. Meyn, R. L. Tweedie, and P. W. Glynn, *Markov Chains and Stochastic Stability*, 2nd ed., ser. Cambridge Mathematical Library. Cambridge, UK: Cambridge University Press, Apr. 2009.

[47] 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.

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

[49] G. Rovatsos, G. V. Moustakides, and V. V. Veeravalli, "Quickest detection of a dynamic anomaly in a sensor network," in *Proc. IEEE 53nd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Nov. 2019.

[50] 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 (ISIT)*, Los Angeles, USA, Jun. 2020.

[51] G. Rovatsos, G. V. Moustakides, and V. V. Veeravalli, "Quickest detection of moving anomalies in sensor networks," *arXiv e-prints*, p. arXiv:2007.14475, Jul. 2020.

[52] 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.

[53] G. Rovatsos, V. V. Veeravalli, D. Towsley, and A. Swami, "Quickest detection of anomalies of varying location and size in sensor networks," *IEEE Trans. Aero. Elec. Sys. (to appear).*

[54] M. Barni and B. Tondi, "Binary hypothesis testing game with training data," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4848–4866, Aug. 2014.

[55] M. Barni and B. Tondi, "Multiple-observation hypothesis testing under adversarial conditions," in *Proc. IEEE Inter. Workshop on Inform. Forensics and Security (WIFS)*, Guangzhou, China, Nov. 2013, pp. 91–96.

[56] M. Barni and B. Tondi, "Adversarial source identification game with corrupted training," *IEEE Trans. Inform. Theory*, vol. 64, no. 5, pp. 3894–3915, Feb. 2018.

[57] Y. Jin and L. Lai, "Adversarially robust hypothesis testing," *in Proc. IEEE 53rd Asilomar Conf. on Signals, Systems, and Computers*, pp. 1806–1810, Nov. 2019.

[58] R. Zhang and S. Zou, "A game-theoretic approach to sequential detection in adversarial environments," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Los Angeles, USA, Jun. 2020, pp. 1153–1158.

[59] S. Stankovic, N. Ilic, M. Stankovic, and K. Johansson, "Distributed change detection based on a consensus algorithm," *IEEE Trans. Signal Proc.*, vol. 59, no. 12, pp. 5686–5697, Dec. 2011.

[60] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in minimax settings," *IEEE Trans. Inform. Theory*, vol. 59, no. 10, pp. 6917–6931, Jul. 2013.

[61] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection with on-off observation control," *Sequential Analysis*, vol. 31, no. 1, pp. 40–77, Feb. 2012.

[62] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in sensor networks," *IEEE Trans. Signal Proc.*, vol. 63, no. 14, pp. 3727–3735, May 2015.

[63] 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.