Quickest Change Detection in the Presence of Covert Adversaries

Amir Reza Ramtin, James Z Hare, Lance Kaplan, Philippe Nain, Venugopal Veeravalli, Don Towsley

Quickest Change Detection in the Presence of Covert Adversaries

A. Ramtin, Z. Hare, L. Kaplan, P. Nain, V. Veeravalli, D. Towsley

***Abstract*—This paper investigates the damage that an adversary can effect while remaining covert in the presence of the Cumulative Sum (CuSum) procedure. An adversary is covert if the time to detection is on the same order as the time to false alarm. Damage is given as an increasing function of the KL-divergence of the adversarial actions and the normal distribution prior to the adversarial attack. By analyzing the problem with a focus on the growth function $g(n)$, which measures the cumulative expected log-likelihood ratio after n time slots following the change, we establish conditions under which the adversary remains covert and provide an analysis of the impact of different adversarial strategies on damage.**

***Index Terms*—CuSum, Asymptotic Analysis, Covertness, Time-Varying Distributions.**

I. Introduction

As technology evolves and the interconnectedness of devices increases, particularly in military and battlefield scenarios, the need for timely and accurate detection of system anomalies becomes increasingly critical. Sequential change detection is a pivotal technique in this context, enabling the prompt identification of moments when the statistical properties of a monitored process undergo significant changes. This timely detection is essential for initiating swift responses and mitigating potential damages in sensitive and high-stakes environments.

However, the effectiveness of sequential change detection can be severely undermined by covert (undetectable) adversarial actions. Adversaries employing covertness strategies can strategically obscure their activities, making it challenging for traditional detection methods to distinguish between normal and malicious behaviors. To address this challenge, it is important to explore the asymptotic limits of covertness—specifically, how subtly an adversary must behave to remain undetected over time. These limits provide deep insights into the potential vulnerabilities of detection systems and the extent to which adversaries can operate undetected. Note that we consider a concept of damage (or reward) that adversaries aim to maximize while remaining covert.

The concept of covertness was first introduced in [1], where the asymptotic limits under which communication could remain undetectable in an AWGN channel were explored. In these scenarios, signals are typically zero-mean, and the challenge lies in detecting subtle changes in power or other signal parameters. Since then, numerous studies have expanded on this idea, focusing on covert communication strategies across various domains [2]. Additionally, [3] has explored the fundamental limits of covert DDoS attacks, a different class of threat where both the mean and variance of network traffic are non-zero, and the adversary manipulates these parameters to avoid detection.

In the context of sequential change detection, we define an adversary as covert if it can strategize to make the average time to detection asymptotically of the same order to the average time to false alarm. This scenario effectively renders the detector ineffective, as it cannot reliably distinguish between normal and malicious conditions.

In our study, we focus on the adversaries employing non-stationary post-change distributions as a strategy to maximize their damage while remaining covert against optimal detectors. It is important to note that, under certain conditions, the Cumulative Sum (CuSum) procedure [4] is recognized as optimal for sequential change detection when addressing non-stationary post-change distributions [5].

A critical parameter in the CuSum procedure is γ, which specifies the minimum average time to false alarm that the procedure is designed to achieve. In [6], the limits of covert communication are studied under the assumption that γ is known and the post-change distributions are stationary. In contrast, we assume that the adversary does not know γ but has control over either the post-change distributions or the timing of their actions, aiming to maximize the damage inflicted. This setting presents a unique challenge, as the adversary can adjust their strategy and vary their attack patterns.

We explore the extent of damage that an adversary can achieve while remaining covert.

The remainder of this paper is organized as follows. In Section II, we provide the background necessary to understand the CuSum procedure and its application to non-stationary post-change distributions. Section III presents the formal problem definition, introducing damage and the concept of covertness and two adversarial strategies studies in the remainder of the paper. This material is presented in the context of quickest change detection. Section IV, contains our main asymptotic results. In Section V we investigate the maximum damage and the conditions under which the adversary can remain covert, across different scenarios. Section VI presents the numerical results and evaluation. Finally, we conclude the paper with a

A. Ramtin and D. Towsley are in College of Information & Computer Sciences, UMass, Amherst, USA. Email: {aramtin, towsley}@umass.edu; P. Nain is at INRIA, Sophia-Antipolis, France. Email: philippe.nain@inria.fr; Z. Hare and L. Kaplan are at DEVCOM Army Research Laboratory, Adelphi, USA. Email: {james.z.hare.civ, lance.m.kaplan.civ}@army.mil; V. Veeravalli is at U. Illinois, Urbana-Champaign, IL. Email: vvv@illinois.edu.

This research is supported by the DEVCOM Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196 (IoBT CRA) and the National Science Foundation under Grant ECCS-2148159.

summary of our findings.

II. Background

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables, and let ν be an unknown but deterministic change-point. Assume that $X_1, \ldots, X_{\nu-1}$ all have density p_0, and that $X_\nu, X_{\nu+1}, \ldots$ have densities $p_{1,1}, p_{1,2}, \ldots$ respectively, with respect to a common dominating measure. Observations are allowed to be non-stationary after the change-point. We denote by $\mathbb{P}_\nu$ the probability measure on the entire sequence of observations when the change-point is ν. That is, under $\mathbb{P}_\nu$ the random variables $X_1, \ldots, X_{\nu-1}$ are i.i.d. with common (pre-change) density p_0, and $X_\nu, X_{\nu+1}, \ldots$ are independent with (post-change) densities $p_{1,1}, p_{1,2}, \ldots$. Let $\mathbb{E}_\nu$ denote the corresponding expectation. For $\nu = \infty$, i.e., no change has occured, this distribution is denoted by $\mathbb{P}_\infty$ and the corresponding expectation by $\mathbb{E}_\infty$.

Note that we assume the post-change distribution is invariant to the change-point ν [5]. Let τ be a stopping time defined on the observation sequence associated with a stopping rule. This means τ is the time at which we stop collecting observations and declare that the change has occurred.

The objective of the quickest change detection (QCD) problem is to develop a stopping rule that minimizes detection delay subject to a false alarm constraint. The traditional approach to achieve the objective is to solve the following optimization problem.

$$\text{ADD}(\gamma) := \inf_{\tau \in \mathcal{C}_\gamma} \sup_{\nu \geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu[(\tau - \nu) + 1)^+]$$

where the expression within the infimum characterizes the worst-case expected delay, denoted by WADD(τ), and ess sup stands for essential supremum. The false alarm constraint set is

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

which guarantees that average time to the false alarm exceeds γ.

For $i \geq \nu$, introduce

$$Z_{i,\nu} = \log \frac{p_{1,i-\nu+1}(X_i)}{p_0(X_i)}.$$

By definition of $\mathbb{P}_\nu$, note that $\mathbb{E}_\nu[Z_{i,\nu}]$ is the KL-divergence of $p_{1,i-\nu+1}$ from p_0. This quantity is always non-negative. However, we make the more restrictive assumption that $\mathbb{E}_\nu[Z_{i,\nu}] > 0$ for all $i \geq \nu$.

Let $g_\nu : \mathbb{R}^+ \to \mathbb{R}^+$ be a strictly increasing and continuous function, which we will refer to as *growth function* as defined in [5]. Note that the inverse of g_ν, denoted by g_ν^{-1} exists and is a strictly increasing and continuous function. We assume that

$$g_\nu(n) = \sum_{i=\nu}^{n+\nu-1} \mathbb{E}_\nu[Z_{i,\nu}], \quad n = 1, 2, \ldots,$$

namely, the cumulative KL-divergence matches the value of the growth function at all positive integers.

We assume that

$$g^{-1}(x) := \sup_{\nu \geq 1} g_\nu^{-1}(x)$$

exists for all $x > 0$. Notice that g^{-1} is also strictly increasing and continuous. We also assume that $\lim_n g_\nu(n) = \infty$ for all $\nu \geq 1$, so that $g^{-1}(x)$ is properly defined on the entire positive real line. Because the post-change distribution is invariant to the change-point ν, we have $g \equiv g_\nu$ and $g^{-1} \equiv g_\nu^{-1}$ for all $\nu \geq 1$.

We define the CuSum stopping rule as

$$\tau(\gamma) := \inf\big\{n : \max_{1 \leq k \leq n} \sum_{i=k}^{n} Z_{i,k} \geq h\big\}, \tag{1}$$

where $h = \log \gamma$. Then, according to [5], under the assumption that

$$g(n) = \omega(\log n), \tag{2}$$

where $\omega(\log n)$ denotes the set of functions that grow faster than $\log n$ as $n \to \infty$, (11) in [5] holds for $Z_{i,\nu}$ when $i \geq \nu \geq 1$. Furthermore, if conditions (14) and (15) in [5] are also satisfied for $Z_{i,\nu}$, then as $\gamma \to \infty$, it follows that

$$\text{ADD}(\gamma) \sim \text{WADD}(\gamma) \sim g^{-1}(\log \gamma), \tag{3}$$

where the symbol "$\sim$" denotes asymptotic equivalence.

The results in [5] are crucial for our analysis of the CuSum procedure in non-stationary environments when conditions (11), (14), and (15) in [5] hold because under those conditions the CuSum procedure in (1) is asymptotically optimal.

III. Problem Definition and Approach Overview

We focus on a problem viewed by an adversary who lacks knowledge of γ but has the ability to either manipulate the post-change distributions or control the timing of their actions. Their goal is to effect the greatest amount of damage.

By "manipulating the post-change distributions," we mean that the adversary can adjust the parameters of the distribution over time. For instance, in the context of a DDoS attack, the adversary might gradually decrease the rate of attack traffic following the change-point ν, such that $\mathbb{E}[Y_{i+1}] < \mathbb{E}[Y_i]$ for $i \geq 1$, where the random variable Y_i denotes the attack traffic at time $\nu + i - 1$.

On the other hand, by "controlling the timing of their actions," we refer to the adversary's ability to decide whether to act at time $\nu+i-1$. In this scenario, their action modifies the distribution, making the post-change distribution $p_{1,i}$ different from the pre-change distribution p_0. For example, in the DDoS attack scenario, the adversary might maintain a constant mean attack traffic μ_1 if it launches an attack, but choose to launch it at time $\nu+i-1$ with a probability that decreases over time.

Let $\mathcal{D}(p_{1,i} \| p_0)$ denote the Kullback-Leibler divergence (KL-divergence) between the distributions $p_{1,i}$ and p_0. Define cumulative damage function $D(n)$ as

$$D(n) = \sum_{i=1}^{n} d\big(\mathcal{D}(p_{1,i} \| p_0)\big), \tag{4}$$

where $d : \mathbb{R}^+ \to \mathbb{R}^+$ maps the KL-divergence at time $\nu + i - 1$ into a measure of damage. We assume that d is a non-decreasing function of its argument.

Here, we provide three examples of a damage function. Assume $p_0 \sim \mathcal{N}(0,1)$ and $p_{1,i} \sim \mathcal{N}(\mu_i, 1+\sigma_i^2)$. The KL-divergence between $p_{1,i}$ and p_0 is given by $\mathcal{D}(p_{1,i}\|p_0) = \frac{1}{2}\left(\sigma_i^2 + \mu_i^2 - \log(1+\sigma_i^2)\right)$.

1) Let $\mu_i > 0$ and $\sigma_i^2 = 0$. Define the damage function as $d(x) = \sqrt{2x}$. Then, the cumulative damage is $D(n) = \sum_{i=1}^{n} \mu_i$.
2) Let $\mu_i = 0$. Define the damage function as $d(x) = 2(x + \log(1+\sigma_i^2) - \frac{1}{2}\sigma_i^2)$. Then, $D(n) = \sum_{i=1}^{n} \log(1+\sigma_i^2)$.
3) Let $\mu_i > 0$ and $\sigma_i^2 > 0$. Define the damage function as $d(x) = 2x + \log(1+\sigma_i^2) - \sigma_i^2$. Then, $D(n) = \sum_{i=1}^{n} \mu_i$.

We consider the *total damage* to be $D\big(\text{ADD}(\gamma)\big)$, representing the total damage that the adversary can inflict before being detected under optimal detection. To determine the greatest amount of $D\big(\text{ADD}(\gamma)\big)$, we first analyze the non-stationary CuSum procedure with $h(\gamma) = \log\gamma$ when conditions (11), (14), and (15) from [5] are met.

Note that if $g(n) = \mathcal{O}(\log n)$, we cannot use (3) in our analysis because condition (2) is not met. In this situation, the CuSum procedure with $h(\gamma) = \log\gamma$ may not be the optimal detector.

We now define *covertness* for the problem of quickest change detection as follows. This definition is analogous to the concept of covert communication against sequential change-point detection as described in Equation (3) of [6].

Definition 1. An adversary is covert if the asymptotic average detection delay,

$$\text{ADD}(\gamma) \sim c\gamma \tag{5}$$

for some $0 < c \leq 1$.

Note that c cannot exceed one because $\text{ADD}(\gamma) \leq \gamma$. This is due to the drift of the log-likelihood ratio being positive after the change, leading to detection, and negative before the change, with γ specifying the minimum time to a false alarm.

Interestingly, considering the covertness condition (5) and applying (3), we derive $g(n) = \log n + \mathcal{O}(1)$. This indicates that when (2) (i.e., condition (11) in [5]) is satisfied, covertness does not hold, and vice versa.

Recall that we aim to analyze the maximum amount of damage while maintaining covertness. Therefore, to advance our analysis, we must address the potential non-optimality of the non-stationary CuSum procedure under the condition $g(n) = \mathcal{O}(\log n)$.

In Section IV, using the continuity property of the function $g(n)$, we show that under optimal detection, as $\gamma \to \infty$, when $g(n) = c\log n$ with $0 < c \leq 1$ or when $g(n) = o(\log n)$, we have $\text{ADD}(\gamma) = \Theta(\gamma)$, where $\Theta(\gamma)$ denotes the set of functions that grow asymptotically at the same rate as γ. Note that this also satisfies (5) (the covertness condition).

In Section V, we explore the amount of damage under optimal detection in three different scenarios $g(n) = \omega(\log n)$, $g(n) = c\log n$ with $0 < c \leq 1$, and $g(n) = o(\log n)$, ensuring that $\mathbb{E}_\infty[\tau] = \gamma$. In particular, we define

$$\begin{aligned} D^* = \; & D\big(\text{ADD}(\gamma)\big) \\ & \text{subject to } g(n) = \log n, \end{aligned} \tag{6}$$

and investigate if D^* is the maximum damage or not.

IV. Analysis of $\text{ADD}(\gamma)$ when $g(n) = \mathcal{O}(\log n)$

To facilitate our analysis, we introduce a parameter δ, allowing the order of $g(n)$ to change with different values of δ. We assume the growth function $g(n)$ is dependent on and continuous with respect to a variable δ.

Let conditions (11), (14), and (15) in [5] hold when $\delta < \delta_0$. This implies, when $\delta < \delta_0$, given

$$\mathbb{E}_\infty[\tau] = \gamma,$$

we have

$$\text{ADD}(\gamma, \delta) = g^{-1}(\log\gamma, \delta),$$

where $g^{-1}(x, \delta)$ is the inverse of $g(n, \delta)$.

Now, assume that $g(n, \delta) = \log n$ when $\delta = \delta_0$, and that

$$\lim_{n\to\infty} \frac{g(n,\delta)}{\log n} < 1,$$

i.e., $g(n, \delta) = c\log n$ with $0 < c < 1$ or $g(n, \delta) = o(\log n)$, when $\delta > \delta_0$. We have

$$\begin{aligned} \lim_{\delta\to\delta_0^-} \text{ADD}(\gamma, \delta) &= \lim_{\delta\to\delta_0^-} g^{-1}(\log\gamma, \delta) \\ &= \Theta(\gamma). \end{aligned}$$

Suppose there exists an optimal detector $\Delta(h)$ that satisfies $\mathbb{E}_\infty[\tau] = \gamma$ for $\delta \geq \delta_0$. Note that $g(n, \delta)$ grows slower for $\delta \geq \delta_0$ compared to $\delta < \delta_0$. Hence, we can infer that when $\delta \geq \delta_0$, the average detection time under this optimal detector is higher than that under a CuSum procedure with $h = \log\gamma$ and $\delta < \delta_0$. This implies, when $\delta < \delta_0$, $\text{ADD}(\gamma, \delta) \geq c\gamma$ with $0 < c \leq 1$.

On the other hand, since under any optimal detector, $\text{ADD}(\gamma, \delta) \leq \mathbb{E}_\infty[\tau]$, given $\mathbb{E}_\infty[\tau] = \gamma$, we have that $\text{ADD}(\gamma, \delta) \leq \gamma$. Therefore, under $\Delta(h)$ when $\delta \geq \delta_0$, we have $\text{ADD}(\gamma, \delta) = \Theta(\gamma)$.

This leads to the conclusion that under any optimal detector, when $\lim_{n\to\infty} \frac{g(n)}{\log n} \leq 1$, we have

$$\text{ADD}(\gamma, \delta) = \Theta(\gamma).$$

which, according to Definition 1, suffices for covertness.

Note that under the condition $g(n) = c\log n$ with $c > 1$, although the optimality of the CuSum procedure with $h = \log\gamma$ has not been rigorously verified, it results in $\mathbb{E}_\infty[\tau] = \Theta(\gamma)$ and $\text{ADD}(\gamma) = \Theta(\sqrt[c]{\gamma})$, which suggests a form of optimality, as the average time to a false alarm remains linear in γ. In contrast, for $0 < c < 1$ or $g(n) = o(\log n)$, the CuSum procedure with $h = \log\gamma$ leads to an average time to a false alarm growing faster than γ, i.e., $\mathbb{E}_\infty[\tau] = \omega(\gamma)$.

V. Analysis of Adversarial Damage

In this section, we analyze the amount of damage an adversary can inflict while remaining covert. We assume that the adversary lacks knowledge of γ, but they can either manipulate the parameters of the post-change distributions, probabilistically choose whether to take action, or deterministically time their actions as time progresses. In all scenarios, the pre-change distributions are standard normal, and the total damage is defined as the cumulative difference between the means of the post-change and pre-change distributions up to the point of detection.

Due to the underlying assumption that the post-change distribution is invariant to the change-point ν, unless explicitly specified, we simply assume $\nu = 1$ or $\nu = \infty$.

A. Gaussian Distribution with Decaying Mean

Consider the following pre-change and post-change distributions,

- pre-change: $p_0 \sim \mathcal{N}(0, 1)$,
- post-change: $p_{1,i} \sim \mathcal{N}(i^{-\delta}, 1),\ 0 < \delta < 1$.

In this section, we demonstrate that the adversary can achieve a maximum total damage of $\Theta(\sqrt{\gamma})$, particularly when $\delta = 0.5$.

The log-likelihood ratio is

$$Z_i = \log \frac{p_{1,i}(X_i)}{p_0(X_i)} = i^{-\delta} X_i - \frac{i^{-2\delta}}{2}.$$

and as a result we have

$$\mathcal{D}(p_{1,i}||p_0) = \mathbb{E}_1\left[Z_i\right] = i^{-\delta}\mathbb{E}_1\left[X_i\right] - \frac{i^{-2\delta}}{2} = \frac{i^{-2\delta}}{2}.$$

which implies that the growth function is

$$g(n) = \sum_{i=1}^{n} \frac{i^{-2\delta}}{2}.$$

In the following analysis, we investigate the extent of damage for $g(n) = \omega(\log n)$, $g(n) = \Theta(\log n)$, and $g(n) = o(\log n)$.

Our first objective is to assess the damage extent when $g(n) = \omega(\log n)$, implying that $0 < \delta < 0.5$.

We know that $g(n)$ is a similar form to the Riemann zeta function. Consequently, we use the Euler-Maclaurin summation formula to relate the asymptotic behavior of $g(n)$ to an integral. It follows that

$$\begin{aligned} g(n) &\approx \frac{1}{2}\int_1^n x^{-2\delta}\, dx \\ &= \frac{n^{1-2\delta} - 1}{2(1-2\delta)}. \end{aligned}$$

For large values of x, we can approximate the inverse function $g^{-1}(x)$ as

$$g^{-1}(x) \approx \left(2(1-2\delta)x + 1\right)^{\frac{1}{1-2\delta}}.$$

We can demonstrate the satisfaction of conditions (13), (16) and (17) as stated in [5]. Hence,

$$\mathrm{ADD}(\delta, \gamma) \approx \left(1 + 2(1-2\delta)\log\gamma\right)^{\frac{1}{1-2\delta}}.$$

Note that the cumulative damage, as defined in (4), with the choice of $d(x) = \sqrt{2x}$ (analogous to Example 1 in Section III), is given by $D(n) = \sum_{i=1}^{n} i^{-\delta}$. Therefore, given $0 < \delta < 0.5$, the total damage, defined as $D\big(\mathrm{ADD}(\delta, \gamma)\big)$, is

$$\begin{aligned} D(\delta, \gamma) &= \sum_{i=1}^{\mathrm{ADD}(\delta,\gamma)} i^{-\delta} \\ &\approx \frac{x^{1-\delta}}{1-\delta}\bigg|_1^{\mathrm{ADD}(\delta,\gamma)} = \frac{\left(1 + 2(1-2\delta)\log\gamma\right)^{\frac{1-\delta}{1-2\delta}} - 1}{1-\delta}. \end{aligned}$$

We can verify that $dD/d\delta$ is positive when $0 < \delta < 0.5$ and thus the total damage increases in δ for $0 < \delta < 0.5$.

Next, we proceed to investigate the amount of total damage under the case $\delta = 0.5$.

First, choose $h(\gamma) = \frac{1}{2}\log\gamma$. We have

$$\lim_{\delta \to 0.5^-} \mathrm{ADD}(\delta, \sqrt{\gamma}) = e^{2h(\gamma)} = \gamma. \tag{7}$$

This implies that the average detection delay converges to γ as δ approaches 0.5 when $0 < \delta < 0.5$ with a threshold less than $\log\gamma$.

Now, consider an optimal detector that satisfies $\mathbb{E}_\infty[\tau] \geq \gamma$ when $\delta = 0.5$. Note that the average detection delay under this detector is higher than it under the optimal detector (the CuSum procedure with $h(\gamma) = \log\gamma$) with $0 < \delta < 0.5$. Furthermore, note that the average detection delay of the CuSum procedure decreases as its threshold decreases. Subsequently, following (7) and noting that $\mathrm{ADD}(\delta, \gamma) \leq \gamma$, we conclude

$$\mathrm{ADD}(0.5, \gamma) = \Theta(\gamma).$$

It follows that

$$D^* = \sum_{i=1}^{\Theta(\gamma)} i^{-0.5} \sim \Theta(\sqrt{\gamma}).$$

We can show that when $\delta > 0.5$, we have $\mathrm{ADD}(\delta) = \gamma$, and thus

$$D(\delta, \gamma) \leq \sum_{i=1}^{\gamma} i^{-\delta} = o(\sqrt{\gamma}).$$

Therefore, we conclude that $\Theta(\sqrt{\gamma})$ is the maximum amount of damage that the adversary can achieve, which occurs when $\delta = 0.5$.

B. Fixed Gaussian Distribution with Decaying Action Probability

In some scenarios, an adversary may not have the ability to arbitrarily change the distribution but can only turn it on or off. This leads to the question: What if the adversary can only decide whether to take an action at a given time?

Assuming a pre-change distribution $p_0(x)$, we further extend our analysis to incorporate an adversary who makes decisions regarding whether to take an action at a given time. In this scenario, the probability of the adversary taking an action is time-dependent and decays over time. In this context, each observation is independently sampled from the distribution $q(x)$ with a probability of occurrence determined

by $r(i)$. Conversely, with a complementary probability of $\bar{r}(i) = 1 - r(i)$, the observations are drawn from the pre-change distribution $p_0(x)$. Consequently, the post-change distribution $p_{1,i}(x)$ can be expressed as a weighted combination: $p_{1,i}(x) = r(i)q(x) + \bar{r}(i)p_0(x)$.

Our initial step involves deriving an expression for the KL-divergence between $p_{1,i}(x)$ and $p_0(x)$ as a function of $p_0(x)$, $q(x)$, $r(i)$, and a term which is bounded by $\mathcal{D}(q||p_0)$. Subsequently, we introduce a decaying function $r(i) = i^{-\delta}$ such that $0 < \delta < 1$ and proceed with our analysis. Finally, we demonstrate that the adversary can achieve a maximum damage of $\Theta(\sqrt{\gamma})$, which occurs when $\delta = 0.5$.

The KL-divergence between two distributions $p_{1,i}$ and p_0 is

$$\begin{aligned}
&\mathcal{D}(p_{1,i}\|p_0)\\
&= \int_{-\infty}^{\infty} \log(\frac{p_{1,i}(x)}{p_0(x)}) p_{1,i}(x)dx,\\
&= \int_{-\infty}^{\infty} \log(\frac{r(i)q(x) + \bar{r}(i)p_0(x)}{p_0(x)}) \big(r(i)q(x) + \bar{r}(i)p_0(x)\big)dx,\\
&= \int_{-\infty}^{\infty} \log(1 + r(i)\big(\frac{q(x)}{p_0(x)} - 1\big)) \big(r(i)q(x) + \bar{r}(i)p_0(x)\big)dx.
\end{aligned}$$

which using the Taylor expansion of $\log(1+x)$, following the detailed computations, can be written as

$$\mathcal{D}(p_{1,i}\|p_0) = r^2(i)\big(\int \frac{q^2(x)}{2p_0(x)}dx - \frac{1}{2}\big) + r^3(i)R(i),$$

where $R(i)$ represents the summation of terms with the factor of $r(i)$ to the power of three or more.

We observe that $\mathcal{D}\left(p_{1,i}\|p_0\right)$ is maximized when $r(i) = 1$, indicating that $\mathcal{D}\left(p_{1,i}\|p_0\right) \leq \mathcal{D}\left(q\|p_0\right)$. Note that both $\mathcal{D}\left(p_{1,i}\|p_0\right)$ and $\mathcal{D}\left(q\|p_0\right)$ are finite since the support of $p_{1,i}$ and q is contained within the support of p_0.

Without loss of generality, we assume $r(1) = 1$. Then, we have

$$\begin{aligned}
\mathcal{D}\left(p_{1,1}\|p_0\right) &= \mathcal{D}\left(q\|p_0\right)\\
&= \int \frac{q^2(x)}{2p_0(x)}dx - \frac{1}{2} + R(1).
\end{aligned}$$

Applying the weighted Jensen's inequality, we can show that

$$\int_{-\infty}^{\infty} \frac{q^2(x)}{p(x)}dx \geq 1.$$

This implies that

$$\int \frac{q(x)^2}{2p_0(x)}dx - \frac{1}{2} \geq 0$$

and thus

$$R(1) \leq \mathcal{D}\left(q\|p_0\right).$$

Furthermore, for all $i \in \mathbb{N}^+$, we have $R(i) \leq R(1)$ because $r(i) \leq r(1)$. Therefore, we conclude that

$$R(i) \leq \mathcal{D}(q\|p_0), \quad i \in \mathbb{N}^+.$$

Now, we consider the scenario where $r(i) = i^{-\delta}$, where $0 < \delta < 1$. In this case, the decay function $r(i)$ follows a power-law decay with the exponent δ. We also assume that the pre-change distribution follows a standard normal distribution, $p_0 \sim \mathcal{N}(0,1)$, and the post-change distribution is represented by $p_{1,i}(x) = r(i)q(x) + \bar{r}(i)p_0(x)$, where $q \sim \mathcal{N}(\mu, 1)$ such that μ is the mean of the distribution.

By substituting these values into the expressions derived earlier, we can analyze the properties and behavior of the growth function $g(n)$ in this specific setting. After undertaking the requisite computations, we get

$$\begin{aligned}
g(n) =& \sum_{i=1}^{n} \mathcal{D}(p_{1,i}\|p_0)\\
=& \frac{1}{2}\big(e^{\frac{\mu^2}{2}} - 1\big)\sum_{i=1}^{n} i^{-2\delta} + \sum_{i=1}^{n} i^{-3\delta}R(i),
\end{aligned}$$

where $R(i) \leq \frac{1}{2}\mu^2$ for all $i \in \mathbb{N}$. This implies

$$g(n) = \frac{1}{2}\big(e^{\frac{\mu^2}{2}} - 1\big)\sum_{i=1}^{n} i^{-2\delta} + \mathcal{O}\big(\frac{1}{2}\mu^2\sum_{i=1}^{n} i^{-3\delta}\big)$$

which implies

$$g(n) \sim \begin{cases} \Theta(n^{1-2\delta}), & \text{if } \delta \in (0, 0.5) \cup (0.5, 1),\\ \Theta(\log n), & \text{if } \delta = 0.5. \end{cases}$$

Now, we consider (6), which conditions the extent of damage under the assumption that $g(n) = \log n$. This assumption is satisfied when $\delta = 0.5$.

Due to the similarity in the form of the growth function to that described in Section V-A, we can use analogous mathematical derivations to establish that the maximum damage the adversary can achieve is $\Theta(\sqrt{\gamma})$, attained when $\delta = 0.5$.

Note that the results in this section, as well as in Section V-A, can be extended as follows.

Assume the pre-change distribution $p_0 \sim \mathcal{N}(0,1)$ and the post-change distributions are given by

$$p_{1,i}(x) = r(i)q(x) + \bar{r}(i)p_0(x),$$

where $q \sim \mathcal{N}\big(\mu(i), 1\big)$. Then, under the constraint $\mathbb{E}_{\infty}[\tau] \geq \gamma$, the maximum damage is $\Theta(\sqrt{\gamma})$, attained when

$$r(i)\mu(i) = 2i^{-0.5}.$$

C. Growing Timing of Actions (Deterministic Action Rules)

We explore the impact of time-dependent action patterns, assuming these patterns are known to the detector. This differs from Section V-B, where the adversary probabilistically chooses their actions.

Assume the adversary controls the timing of its actions in the following manner. Let $f : \mathbb{N} \to \{0, 1\}$. When $f(i) = 1$, the adversary chooses $\mathcal{N}(\mu, 1)$ as the post-change distribution, and if $f(i) = 0$, it chooses $\mathcal{N}(0, 1)$, at time i. The pre-change and post-change distributions are therefore characterized as

- Pre-change: $p_0 \sim \mathcal{N}(0,1)$
- Post-change: $p_{1,i} \sim \begin{cases} \mathcal{N}(\mu, 1), & \text{if } f(i) = 1 \text{ for } i \in \mathbb{N}^+\\ \mathcal{N}(0, 1), & \text{otherwise.} \end{cases}$

Let's $q(x)$ denote a Gaussian distribution with mean μ and variance 1. The growth function is

$$\begin{aligned} g(n) &= \sum_{i=1}^{n} \mathcal{D}(p_{1,i} \| p_0) \\ &= \mathcal{D}(q \| p_0) \sum_{i=1}^{n} \mathbf{1}_{[f(i)=1]} \\ &= \frac{1}{2}\mu^2 \sum_{i=1}^{n} \mathbf{1}_{[f(i)=1]}. \end{aligned}$$

Now, we assume the statement presented in (1) is satisfied. This implies that $g(n) = \sum_{i=1}^{n} \mathbf{1}_{[f(i)=1]}$ must grow logarithmically as n goes infinity. An example of such a condition can be defined as follows: $f(i) = 1$ if $i = \lfloor e^j \rfloor$, where $j \in \mathbb{N}^+$.

This condition on $f(i)$ implies

$$\begin{aligned} D^* &\leq \mu \sum_{i=1}^{\gamma} \mathbf{1}_{[f(i)=1]} \\ &= \Theta(\log \gamma). \end{aligned}$$

Notably, when $g(n) = \omega(\log n)$, conditions (11), (14), and (15) in [5] hold, which, following (3) and the necessary calculations, establishes that the damage achievable by the adversary is $\Theta(\log \gamma)$.

Furthermore, when $g(n) = o(\log n)$, the damage grows at an order of $o(\log \gamma)$.

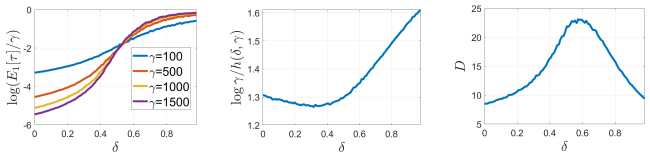
Under these observations, we conclude that $\Theta(\log \gamma)$ is the maximum amount of damage. This implies that time-dependent action patterns, when known to the detector, do not yield any additional reward.

VI. Numerical Results

In this section, we conduct a simulation study to assess the performance of the CuSum algorithm under non-stationary post-change conditions. The pre-change observations follow a standard normal distribution $\mathcal{N}(0, 1)$, while the post-change observations adhere to a normal distribution with a mean parameterized by $\mathcal{N}(\mu_1 \times (1 + i - \nu)^{-\delta}, 1)$, reflecting the non-stationarity introduced after the change-point ν. Each simulation is executed 10^4 times for robustness, with the threshold $h(\delta, \gamma)$ chosen such that

$$|\mathbb{E}_\infty[\tau] - \gamma|/\gamma < 0.005. \tag{8}$$

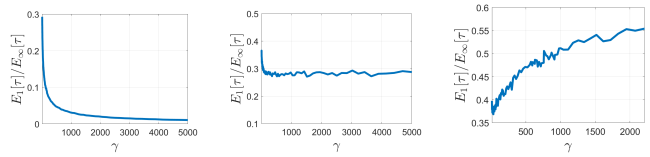
We begin by examining a scenario where $\mu_1 = \sqrt{2}$ and δ varies incrementally from 0 to 1 in steps of 0.02. It is important to note that when $\delta = 0.5$, the function $g(n) = \log n$. Figure 1(a) illustrates the logarithm of the ratio $\mathbb{E}_1[\tau]$ to γ for four distinct values of γ, revealing a phase transition at $\delta = 0.5$. Figure 1(b) depicts the ratio of $\log \gamma$ to $h(\delta, \gamma)$ with $\gamma = 1000$, where a rapid increase is observed when $\delta > 0.5$. This behavior suggests the non-optimality of the CuSum procedure with a threshold of $\log \gamma$ for $\delta > 0.5$ because one must choose $h(\delta, \gamma) = o(\log \gamma)$ for $\delta > 0.5$ to satisfy (8). Finally, Figure 1(c) presents the total damage as a function of δ for $\gamma = 1000$. We expect the point of maximum damage converges to $\delta = 0.5$ as $\gamma \to \infty$.



(a) (b) (c)

Fig. 1. (a) Logarithm of ratio of $\mathbb{E}_1[\tau]$ to γ for four different values of γ, ratio of (b) $\log \gamma$ to $h(\delta, \gamma)$ satisfying (8), and (c) total damage, with $\gamma = 10^3$.

In the second scenario, γ ranges from 10 to at most 10^5, and the evaluation is conducted for three different values of δ, with $\mu_1 = 1$. The results are illustrated in Figure 2. We observe that when $\delta = 0.5$, we have $\mathbb{E}_1[\tau] = \Theta(\mathbb{E}_\infty[\tau])$, i.e., $\mathbb{E}_1[\tau] = \Theta(\gamma)$, and when $\delta > 0.5$, we have $\mathbb{E}_1[\tau] \sim \mathbb{E}_\infty[\tau]$, i.e., $\mathbb{E}_1[\tau] \sim \gamma$.



(a) (b) (c)

Fig. 2. Ratio of $\mathbb{E}_1[\tau]$ to γ with (a) $\delta = 0.25$, (b) $\delta = 0.5$, and (c) $\delta = 0.6$.

VII. Conclusion

We investigated the behavior of the CuSum procedure under non-stationary post-change distributions within the context of quickest change detection. By focusing on the growth function $g(n)$, we explored different adversarial strategies and determined the conditions under which the adversary remains covert as well as they inflict the maximum amount of damage.

References

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," IEEE journal on selected areas in communications, vol. 31, no. 9, pp. 1921–1930, 2013.

[2] X. Chen, J. An, Z. Xiong, C. Xing, N. Zhao, F. R. Yu, and A. Nallanathan, "Covert communications: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1173–1198, 2023.

[3] A. R. Ramtin, P. Nain, D. S. Menasche, D. Towsley, and E. d. S. e Silva, "Fundamental scaling laws of covert ddos attacks," Performance Evaluation, vol. 151, p. 102236, 2021.

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

[5] Y. Liang, A. G. Tartakovsky, and V. V. Veeravalli, "Quickest change detection with non-stationary post-change observations," IEEE Transactions on Information Theory, vol. 69, no. 5, pp. 3400–3414, 2022.

[6] K.-W. Huang, H.-M. Wang, and H. V. Poor, "On covert communication against sequential change-point detection," IEEE Transactions on Information Theory, vol. 67, no. 11, pp. 7285–7303, 2021.